BEST AVAILABLE COPY

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

PCT

(43) International Publication Date
2 December 2004 (02.12.2004)

(10) International Publication Number
WO 2004/104161 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number:
PCT/IL2004/000429

(22) International Filing Date: 20 May 2004 (20.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/441,281   20 May 2003 (20.05.2003)   US

(71) Applicant *(for all designated States except US)*: COMPU-GEN LTD. [IL/IL]; 72 Pinchas Rosen Street, 69 512 Tel Aviv (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LEVANON, Erez [IL/IL]; 73 Menachem Begin Street, 49732 Petach Tikva (IL). POLLOCK, Sarah [IL/IL]; 16 Hoshea Street, Apt. 2, 63 506 Tel Aviv (IL). NEMZER, Sergey [IL/IL]; 5 HaRenanim Street, 52 595 Ramat Gan (IL). SHOSHAN, Avi [IL/IL]; 2/4 Rachavat Har Sinai Street, 82 000 Kiryat Gat (IL). KHOSRAVI, Rami [IL/IL]; 36/1 Halevi Street, 46490 Herzlia (IL). WALACH, Shira [IL/IL]; 40 Bnei Brit Street, 45 265 Hod HaSharon (IL). LEVINE, Zurit [IL/IL]; 4 Reihan Street, 46 419 Herzlia (IL). BERN-STEIN, Jeanne [IL/IL]; 23 Harimon Street, 40 300 Kfar Yona (IL). DAHARY, Dvir [IL/IL]; 23 Pinkas Street, 62 662 Tel Aviv (IL). WASSERMAN, Alon [IL/US]; 564

First Avenue, Apt. 13a, New York, NY 10016 (US). ROT-MAN, Galit [IL/IL]; 5 Yair Shtern Street, 46 412 Herzlia (IL).

(74) Agent: G.E. EHRLICH (1995) LTD.; 11 Menachem Begin Street, 52 521 Ramat Gan (IL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE TRANSCRIPTS AND METHODS, KITS AND ARRAYS UTILIZING SAME

(57) Abstract: A method of identifying putative naturally occurring antisense transcripts is provided. The method is effected by (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database, thereby identifying putative naturally occurring antisense transcripts.

METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE TRANSCRIPTS AND METHODS, KITS AND ARRAYS UTILIZING SAME

BACKGROUND OF THE INVENTION

The present invention relates to the field of naturally occurring antisense transcripts. More particularly, the present invention relates to methods of identifying naturally occurring antisense transcripts, databases storing polynucleotide sequences encoding identified naturally occurring antisense transcripts, oligonucleotides derived therefrom and methods and kits utilizing same.

FIELD OF THE INVENTION

Naturally occurring antisense RNA transcripts are endogenous transcripts, which exhibit complementarity to sense transcripts of which are typically of a known function. It has been established that these endogenous antisense transcripts play an important role in regulating prokaryotic gene expression and are increasingly implicated as involved in eukaryotic gene regulation.

Cis-encoded antisense transcripts are encoded by the same locus as the sense transcripts and are transcribed from strand of DNA opposite to that encoding the sense transcript; as such, cis encoded antisense transcripts are typically completely complementary with a portion of the sense transcript. *Trans*-encoded antisense transcripts are, by contrast, transcripts, which are encoded on a different locus and as such, may display only partial complementarity with a sense transcript.

Natural antisense RNAs were first described in prokaryote studies, which suggested that such transcripts play a role in gene expression regulation. Prokaryotic antisense transcripts are widely distributed and are involved in the control of numerous biological functions including transposition, plasmid replication, incompatibility and conjugation. In prokaryotes, antisense transcripts are typically involved in down-regulation of sense transcript expression, although involvement in positive regulation was also suggested [reviewed in Wagner EG. and Simons RW. (1994) Annu. Rev. Microbiol. 48:713-742].

The first example of transcription from both strands of eukaryotic DNA was illustrated in human and mouse mitochondrial genes [Anderson S. et al. (1981) Nature

2

290:457-465 and Bibb MJ. et al. (1981) Cell 26:167-180]. Since then, examples of antisense transcripts have been documented in a variety of organisms including viruses, slime molds, insects, amphibians and birds as well as mammals. It is thought that these antisense RNAs are involved in extremely diverse biological functions,

5   such as, hormonal response, control of proliferation, development, structure, viral replication and others. Some antisense RNAs are conserved between species suggesting that these antisense RNAs are not fortuitous but rather play an important role in gene expression regulation [Kidny MS. et al. (1987) Mol. Cell Biol. 7:2857-2862, Nepveu A. and Marcu KB. (1986) EMBO J. 5:2859-2865 and Bentley DL. et

10   al. (1986) Nature 321:702-706].

       Antisense transcripts can also encode proteins. Examples for protein encoding antisense transcripts include *rev-ErbAx* [Lazar MA. (1989) Mol. Cell. Biol. 9:1128-1136], *gfg* [Kimelman D. et al. (1989) Cell 59:687-696] and *n-cym* [Armstrong BC. et al. (1992) Cell Growth Differ. 3:385-390]. Such antisense transcripts typically

15   include a distinct open reading frame (ORF) and polyadenylation signal for cytoplasm transportation.

       However, it is believed that most antisense transcripts play a role in gene expression regulation. This assumption is mostly based on spatial and/or temporal distributions of sense and antisense transcripts. Indeed, tissue distribution studies

20   suggest that high levels of sense and antisense transcripts rarely occur together, as was exemplified for the *dopa decarboxylase* transcripts in *Drosophila* [Spencer CA. et al. (1986) Nature 322:279-281]. Additional studies demonstrated that changes in sense gene expression correlate with presence of antisense RNA. Furthermore, an inverse relationship between levels of accumulation of sense and antisense transcripts

25   such as has been reported for *α1 (I) collagen* transcripts in chondrocytes under chemotherapy has also been reported [Farrell CM. And Lukens LN. (1995) J. Biol. Chem. 270:3400-3408]. However, it will be appreciated that mutual expression of sense and their corresponding antisense transcripts is also reported and may involve a different mechanism of regulation.

30   Evidence for involvement of antisense-mediated gene regulation in the development of pathologies has also been presented. For example, endogenous antisense transcripts may be involved in regulation of the expression levels of the

3

tumor suppressor gene WT1 observed in Wilm's tumors [Eccles MR. et al. (1994) Oncogene 9:2059-2063].

Natural antisense regulation of gene expression can be effected via one of several mechanisms.

Nuclear regulation

Nuclear regulation can be effected via several gene-processing pathways [reviewed in Vanhee-Brosollet C. and Vaquero C. (1998) Gene 211:1-9]

dsRNA-mediated DNA methylation - complementation between endogenous sense transcripts and antisense transcripts of sequences as short as 30 bp may initiate DNA-methylation, a well-established phenomenon in a number of organisms [Sharp A. (2001) Genes Dev. 15:485-490]. Methylation can be directed to different portions of an encoding region of the gene or to the promoter region. DNA methylation results in complete suppression of transcription probably by recruitment of histone deacetylases.

Transcriptional regulation – in which case antisense transcription hampers sense transcription. Such interference may involve the collision of two transcription complexes. Alternatively, interference may result from competition on an essential rate limiting transcription factor resulting in premature termination or in reduced elongation of transcription, the transcripts with the highest rate of transcription being predominant.

Post-transcriptional nuclear regulation – involves antisense intervention of either maturation and/or transport of the sense transcript to the cytoplasm. Alternatively, antisense transcripts displaying similar structural features to sense transcripts can bind proteins expected to interact with their sense counterparts, thereby depriving sense messengers from proteins necessary for their function.

Cytoplasmic regulation

Messenger stability –double stranded RNA may affect messenger stability via "RNA interference", which involves short segments of double stranded RNA (dsRNA) homologous in sequence to the silenced gene. These undersized segments, which are generated by a ribonuclease III cleavage of longer dsRNAs, can guide a single stranded target mRNA, via base pairing, to a multisubunit complex which participates in the degradation of the target mRNA. Alternatively, messenger stability

4

may be affected by RNA degradation, which is mediated by double stranded RNA-directed Rnases.

Translation – masking the 3' untranslated region (UTR) and the polyA tail of the sense transcript is believed to modulate translation efficiency probably via direct or indirect interaction between 3'-proximal elements and upstream sequences or structures [reviewed in Jackson RJ. And Standart N. (1990) Cell 62:15-24].

Realizing the fundamental role antisense transcripts play in regulating sense transcription, stability and function, resulted in a number of attempts to systematically identify natural antisense transcripts. Accordingly, differential approaches were taken for exploring non-coding antisense RNA transcripts and antisense transcripts including an ORF. Although the latter carries ORF consensus parameters, uncovering antisense data from general sequence databases has proven to be a complicated task, as many of these sequences include an evolutionary conserved secondary structure rather than a conserved primary sequence, therefore primary sequence alignment methods are often not very effective. Indeed, only a few attempts have been tried to date with only limited success.

Maziel's group [Chen JH. et al. (1990) Comput. Applic. Biosci. 6:7-18 and Le SY. et al (1990) Human Genome Initiative and DNA Recombination Vol. 1:127-136] has experimented with methods that look for regions of a genome with predicted RNA structures that are significantly more stable thermodynamically than random sequence of the same base composition. Although this approach detected a few highly structured non-coding RNAs, as well as few *cis*-regulatory structures, it appears that it is of limited use for large-scale applications.

Another approach examined coding dense genomes, having suspicious-looking large regions with little or no coding potential termed "gray holes" [Olivas WM. et al. (1997) Nucleic acids Res. 25:4619-4625]. Fifty nine gray holes were tested in the yeast genome. Northern analysis detected distinct transcripts from 15 of the gray holes. Only one transcript appeared to be a non-coding antisense transcript illustrating the low efficiency of this method.

SUMMARY OF THE INVENTION

5

The background art does not teach or suggest methods of systematically and efficiently identifying novel naturally occurring antisense molecules and methods of generating and using same for detecting, quantifying and/or regulating sense transcripts, such as for example, mRNA transcripts associated with a pathological state.

According to one aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts, the method comprising: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database according to at least one sequence criterion, thereby identifying putative naturally occurring antisense transcripts, wherein the at least one sequence criterion includes one or more of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

According to another aspect of the present invention there is provided a kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, the sequence region not being complementary with a naturally occurring antisense transcript.

According to yet another aspect of the present invention there is provided a kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA transcript of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest.

According to still another aspect of the present invention there is provided a method of designing artificial antisense transcripts, the method comprising: (a) providing a database of naturally occurring antisense transcripts; (b) extracting from the database criteria governing structure and/or function of the naturally occurring antisense transcripts, wherein the criteria governing structure and/or function of the

6

naturally occurring antisense transcripts include one or more of antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.; and (c) designing the artificial antisense transcripts according to the criteria.

According to further features in preferred embodiments of the invention described below the criteria governing structure and/or function of the naturally occurring antisense transcripts are selected from the group consisting of antisense length, complementarity length, complementarity position, intron molecules,

5   alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and

10   predicted single stranded and double stranded regions.

According to an additional aspect of the present invention there is provided a computer readable storage medium comprising a database including a plurality of sequences, wherein each sequence is of a naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the

15   database further includes information pertaining to each sequence of the naturally occurring antisense transcripts, the information is selected from the group consisting of related sense gene, antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin

20   structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-

binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

According to still further features in the described preferred embodiments the database further includes information pertaining to generation of the database and potential uses of the database.

According to yet an additional aspect of the present invention there is provided a method of generating a database of naturally occurring antisense transcripts, the method comprising: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database so as to identify putative naturally occurring antisense transcripts; and (c) storing sequence information of the identified naturally occurring antisense transcripts, thereby generating the database of the naturally occurring antisense transcripts.

According to still an additional aspect of the present invention there is provided a system for generating a database of a plurality of putative naturally occurring antisense transcripts, the system comprising a processing unit, the processing unit executing a software application configured for: (a) computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and (b) identifying expressed polynucleotide sequences from the second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of the first database according to at least one sequence criterion, wherein the at least one sequence criterion includes at least one of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

According to a further aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts, the method comprising screening a database of expressed polynucleotides sequences according to at least one sequence criterion, the at least one sequence criterion being selected to identify putative naturally occurring antisense transcripts.

8

According to yet a further aspect of the present invention there is provided a method of quantifying at least one mRNA of interest in a biological sample, the method comprising: (a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one mRNA of interest, wherein the at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, the sequence region not being complementary with a naturally occurring antisense transcript; and (b) detecting a level of binding between the at least one mRNA of interest and the at least one oligonucleotide to thereby quantify the at least one mRNA of interest in the biological sample.

According to still a further aspect of the present invention there is provided a method of quantifying the expression potential of at least one mRNA of interest in a biological sample, the method comprising: (a) contacting the biological sample with at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest; and (b) detecting a level of binding between the at least one mRNA of interest and the first oligonucleotide and a level of binding between the naturally occurring antisense transcript complementary to the mRNA of interest and the second oligonucleotide to thereby quantify the expression potential of the at least one mRNA of interest in the biological sample.

According to other aspect of the present invention there is provided a method of quantifying at least one naturally occurring antisense transcript of interest in a biological sample, the method comprising: (a) contacting the biological sample with at least one oligonucleotide capable of binding with the at least one naturally occurring antisense transcript of interest, wherein the at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the naturally occurring antisense transcript of interest, the sequence region not being complementary with a naturally occurring mRNA transcript; and (b) detecting a level of binding between the at least one naturally occurring antisense transcript of interest and the at least one oligonucleotide to thereby quantify the at least one naturally occurring antisense transcript of interest in the biological sample.

9

According to still further features in the described preferred embodiments the first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

According to still further features in the described preferred embodiments the second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

According to still further features in the described preferred embodiments an average sequence length of the expressed polynucleotide sequences of the second database is selected from a range of 0.02 to 0.8 Kb.

According to still further features in the described preferred embodiments the second database is generated by: (i) providing a library of expressed polynucleotides; (ii) obtaining sequence information of the expressed polynucleotides; (iii) computationally selecting at least a portion of the expressed polynucleotides according to at least one sequence criterion; and (iv) storing the sequence information of the at least a portion of the expressed polynucleotides thereby generating the second database.

According to still further features in the described preferred embodiments the at least one sequence criterion for computationally selecting the at least a portion of the expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

According to still further features in the described preferred embodiments the step of testing the putative naturally occurring antisense transcripts for an ability to form the duplex with the at least one sense oriented polynucleotide sequence under physiological conditions.

According to still further features in the described preferred embodiments the method further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the

10

group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

According to still further features in the described preferred embodiments a length of the at least one oligonucleotide is selected from a range of 15-200 nucleotides.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a single stranded oligonucleotide.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a double stranded oligonucleotide.

According to still further features in the described preferred embodiments a guanidine and cytosine content of the at least one oligonucleotide is at least 25 %.

According to still further features in the described preferred embodiments the at least one oligonucleotide is labeled.

According to still further features in the described preferred embodiments the at least one oligonucleotide is attached to a solid substrate.

According to still further features in the described preferred embodiments the solid substrate is configured as a microarray and whereas the at least one oligonucleotide includes a plurality of oligonucleotides each attached to the microarray in a regio-specific manner.

According to still further features in the described preferred embodiments a length of each of the first and second oligonucleotides is selected from a range of 15-200 nucleotides.

According to still further features in the described preferred embodiments the first and second oligonucleotides are single stranded oligonucleotides.

According to still further features in the described preferred embodiments the first and second oligonucleotides are double stranded oligonucleotide.

According to still further features in the described preferred embodiments a guanidine and cytosine content of each of the first and second oligonucleotides is at least 25 %.

According to still further features in the described preferred embodiments the first and second oligonucleotides are labeled.

11

According to still further features in the described preferred embodiments the first and second oligonucleotides are attached to a solid substrate.

According to still further features in the described preferred embodiments the solid substrate is configured as a microarray and whereas each of the first and second oligonucleotides includes a plurality of oligonucleotides each attached to the microarray in a regio-specific manner.

According to yet other aspect of the present invention there is provided a method of identifying a novel drug target, the method comprising: (a) determining expression level of at least one naturally occurring antisense transcript of interest in cells characterized by an abnormal phenotype; and (b) comparing the expression level of the at least one naturally occurring antisense transcript of interest in the cells characterized by an abnormal phenotype to an expression level of the at least one naturally occurring antisense transcript of interest in cells characterized by a normal phenotype, to thereby identify the novel drug target.

According to still further features in the described preferred embodiments the abnormal phenotype of the cells is selected from the group consisting of biochemical phenotype, morphological phenotype and nutritional phenotype.

According to still further features in the described preferred embodiments determining expression level of at least one naturally occurring antisense transcript of interest is effected by at least one oligonucleotide designed and configured so as to be complementary to a sequence region of the at least one naturally occurring antisense transcript of interest, the sequence region not being complementary with a naturally occurring mRNA transcript.

According to still other aspect of the present invention there is provided a method of treating or preventing a disease, condition or syndrome associated with an upregulation of a naturally occurring antisense transcript complementary to a naturally occurring mRNA transcript, the method comprising administering a therapeutically effective amount of an agent for regulating expression of the naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the agent for regulating expression of the naturally occurring antisense transcript is at

12

least one oligonucleotide designed and configured so as to hybridize to a sequence region of the at least one naturally occurring antisense transcript.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a ribozyme.

According to still further features in the described preferred embodiments the at least one oligonucleotide is a sense transcript.

According to a supplementary aspect of the present invention there is provided a method of diagnosing a disease, condition or syndrome associated with a substandard expression ratio of an mRNA of interest over a naturally occurring antisense transcript complementary to the mRNA of interest, the method comprising: (a) quantifying expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest; (b) calculating the expression ratio of the mRNA of interest over the naturally occurring antisense transcript complementary to the mRNA of interest, thereby diagnosing the disease, condition or syndrome.

According to still a supplementary aspect of the present invention there is provided A computer readable storage medium comprising data stored in a retrievable manner, said data including sequence information of co-regulated human polynucleotide sequences as set forth in files seqs_125 and/or seqs_133 of enclosed CD-1, polypeptide sequences encoded by the polynucleotide sequences as set forth in the file pep_seqs_133 of enclosed CD-ROM 3 and sequence annotations as set forth in the file annotations_133 of enclosed CD-ROM 3.

According to still a supplementary aspect of the present invention there is provided a method of modulating an activity or expression of a gene product, the method comprising upregulating or down regulating expression or activity of a naturally occurring antisense transcript of the gene product, thereby modulating the activity or expression of the gene product.

According to still further features in the described preferred embodiments the method further comprising upregulating or down regulating expression or activity of the gene product.

According to still a supplementary aspect of the present invention there is provided an isolated polynucleotide comprising any of the nucleic acid sequences set

forth in the file seqs_125 or seqs_133 of the enclosed CD-ROM1 of the enclosed CD-ROM 1.

According to a supplementary aspect of the present invention there is provided an isolated polypeptide comprising any of the amino acid sequences set forth in the pep_seqs_133 of enclosed CD-ROM-3.

According to still a supplementary aspect of the present invention there is provided a method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a naturally occurring antisense transcript of a gene encoding an expression product targeted by the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent, thereby enhancing the therapeutic efficacy of the therapeutic agent.

According to still a supplementary aspect of the present invention there is provided a method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a naturally occurring antisense transcript of a gene encoding an expression product capable of inhibiting therapeutic activity of the therapeutic agent, prior to, concomitant with, or following provision of the therapeutic agent thereby enhancing the therapeutic efficacy of the therapeutic agent.

According to still a supplementary aspect of the present invention there is provided a method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a polynucleotide complementary to a naturally occurring antisense transcript of a gene encoding an expression product capable of upregulating activity or expression of the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent, thereby enhancing the therapeutic efficacy of the therapeutic agent.

The present invention successfully addresses the shortcomings of the presently known configurations by providing a novel approach for identifying naturally occurring antisense transcripts, methods of designing artificial antisense transcripts according to information derived therefrom, methods and kits using naturally occurring and synthetic antisense transcripts, and by providing the naturally occurring and synthetic antisense transcripts thereof.

14

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable

5   methods and materials are described below.   In case of conflict, the patent specification, including definitions, will control.   In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

10   The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and

15   readily understood description of the principles and conceptual aspects of the invention.   In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

20   In the drawings:

FIG. 1 illustrates EST alignment along genomic DNA, generated according to the teachings of the present invention. Alignment results identify two strand groups of transcripts i.e., sense transcripts and antisense transcripts with an indicated sequence overlap.

25   FIG. 2 is a black box diagram illustrating a system designed and configured for generating a database of naturally occurring antisense sequences generated according to the teachings of the present invention.

FIG. 3 is a black box diagram illustrating a remote configuration of the system described in Figure 2.

30   FIGs. 4a-k are sequence alignments of overlapping regions of selected naturally occurring antisense and sense sequence pairs identified according to the teachings of the present invention.

15

FIGs. 5a-g are sequence alignments of overlapping regions of selected naturally occurring antisense and sense sequence pairs identified according to the teachings of the present invention.

FIG. 6 schematically illustrates two transcription products of 53BP1 gene (red
5   and green) and their corresponding partial complementary antisense transcripts of the 76p gene (blue). Numbers in parenthesis indicate length of sequence complementation. Schematic location of strand-specific RNA probes used for northern blotting of sense (53BP1, Riboprobe#1) and antisense (76p, Riboprobe#2) transcripts is shown.

10 FIG. 7 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of 53BP1 transcripts. Arrows on the right indicate the molecular weight of the identified 53BP1 transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. [Numbers on the left denote the size of molecular weight markers in Kb.

15 FIG. 8 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of 76p transcripts. Arrows on the right indicate the molecular weight of the identified 76p transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. [Numbers on the left denote the size of molecular weight markers in Kb.

20 FIG. 9 is an autoradiogram of a northern blot analysis depicting tissue distribution and expression levels of 76p transcripts. Arrows on the right indicate the molecular weight of the identified 76p transcripts. Numbers on the left denote the migration of molecular weight marker in Kb.

FIG. 10 illustrates the genomic organization of the 53BP1 gene and 76p gene,
25 as elucidated from the RT-PCR analysis presented in the Examples section hereinbelow. Black arrows indicate the location of the primers used for RT-PCR analysis. Asterisks denote stop codons.

FIG. 11 schematically illustrates two transcription products of CIDE-B gene and their corresponding partial complementary antisense transcript of the BLTR2
30 gene. Schematic location of the strand-specific 430 nucleotide RNA probe used for northern analysis of sense (CIDE-B) and antisense (BLTR2) transcripts is shown. Dashed rectangles indicate the predicted coding sequence of the transcripts.

16

FIG. 12 is an autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of BLTR2 transcripts. Arrows on the right indicate the molecular weight of the identified BLTR2 transcripts relative to the migration of 28S and 18S ribosomal RNA subunits. Numbers on the left denote the size of molecular weight markers in Kb.

FIG. 13 shows autoradiogram of a northern blot analysis depicting cellular distribution and expression levels of CIDE-B transcripts. Arrows on the right indicate the molecular weight of the identified CIDE-B transcripts relatively to the migration of 28S and 18S ribosomal RNA subunits. Numbers on the left denote the migration size of molecular weight markers in Kb.

FIG. 14 schematically illustrates a transcription product of APAF-1 gene and its corresponding partial complementary antisense transcripts of the EB-1 gene. Schematic location of the strand-specific 366 nucleotide RNA probe used for northern analysis of sense (APAF-1) and antisense (EB-1) transcripts is shown. Asterisks indicate the predicted coding sequence borders of the transcripts.

FIGs. 15a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of EB-1 (Figure 15a) and APAF-1 transcripts (Figure 15b). Numbers on the left denote the size of molecular weight marker in Kb.

FIG. 16 schematically illustrates a transcription product of the MINK-2 gene and its corresponding partial complementary antisense transcript of the AchR-ϵ gene. Schematic location of the strand-specific 280 nucleotide RNA probe used for northern analysis of sense (Mink-2) and antisense (AchR-ϵ) transcripts is shown.

FIGs. 17a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of AchR-ϵ antisense transcripts (Figure 17a) and the sense complementary transcript of Mink-2 (Figure 17b). Arrows on the right denote the migration of molecular weight markers in Kb.

FIG. 18 schematically illustrates a transcription product of Cyclin-E2 gene and its corresponding partial complementary antisense transcript. Schematic location of strand-specific RNA probes used for northern blotting of sense (Riboprobe#1) and antisense (Riboprobe#2) transcripts is shown.

FIGs. 19a-b are autoradiograms of northern blot analyses depicting cellular distribution and expression levels of Cyclin E2 antisense transcript (Figure 19a) and

the sense complementary transcript (Figure 19b). Arrows on the left denote the migration of molecular weight markers in Kb.

FIG. 20 illustrates results from RT-PCR analysis of the expression patterns of CIDE-B transcript and its complementary naturally occurring antisense transcript following concentration dependent induction of apoptosis. Lanes: (1) 50 μM etoposide; (2) 100 μM etoposide; (3) 250 μM etoposide; (4) 500 μM etoposide; (5) 10 nM staurosporine; (6) 100 nM staurosporine; (7) 250 nM staurosporine; (8) 1000 nM staurosporine; (9) untreated cells (UT).

FIGs. 21a-c are results of RT-PCR analyses depicting expression patterns of AchRϵ and its naturally occurring antisense transcript following time-dependent induction of differentiation. Figure 21a illustrates the position of riboprobes used for reverse transcription reaction. Figure 21b shows the reciprocal expression pattern of sense and antisense transcripts (indicated by arrows). Figure 21c shows the expression pattern of the antisense transcript alone.

FIGs. 22a-j illustrate results of northern blot analysis of sense/antisense clusters revealing positive signals for sense/antisense genes in the microarray analysis. Diagrams describing genomic organization of the relevant region for each of the sense/antisense clusters are included above the autoradiograms, and regions of overlap (including GenBank accession number) from which the strand-specific riboprobes were derived are included. Sense-antisense pair numbers are as they appear in the microarray (as depicted in Table_S2 on the attached CD-ROM2 and in conversion Table 6). Figure 22a reveals expression patterns of randomly selected sequence pair number 235, denoted as Rand_235 in Table 6. Similarly, Figure 22b corresponds to pair number 173, Figure 22c to pair number 248, Figure 22d to pair number 6, Figure 22e to pair number 216, Figure 22f to pair number 239, Figure 22g to pair number 202, Figure 22h to pair number 114, Figure 22i to pair number 188, and Figure 22j to pair number 223. Eight pairs (Figures 22a-h) evaluated revealed positive signals for both sense and antisense expression, while two (Figures 22i-j) revealed a positive signal for only one of the genes, with the counterpart being a known RefSeq mRNA.

FIG. 23 is a Table depicting expression patterns in various cell lines and tissues as probed with a subset of 264 pairs from the putative sense/antisense dataset

18

of the present invention. The pairs are denoted by the pair number and described in Table_S1 of CD-ROM2. "C" and "AC" denote the two counterpart probes. Expression was also verified for positive controls, including the ubiquitously expressed genes *gapdh, actin, hsp70* and *gnb2l1* in various concentrations, and 11 previously documented sense/antisense pairs. Expression thresholds were verified and indicated as "+", if the probe passed the threshold in at least one cell line or tissue or "-", if the probe did not pass the threshold in all experiments. In cases where both the sense and the antisense oligo passed the expression threshold, the antisense was declared "verified". In cases where only one of the probes passed the expression threshold, but the other probe was fully contained within a known mRNA deposited in GenBank, the antisense was declared "indirectly verified". Normalization for microarray signals was conducted as described in the methods section. *Rji* ratios were obtained for each cell line/tissue assessed. Cases of flagged-out spots for which there was no information were marked "−1.00". Data represent values of the two reciprocal experiments.

FIG. 24a is a sequence alignment showing a 5'-5' overlap between Nicastrin (NCSTN, GenBank Accession No. NM_015331) and coatomer protein complex subunit α (COPA, GenBank Accession No. NM_004371).

FIG. 24b is a histogram showing expression of NCSTN and two isoforms of COPA in a number of samples, as determined by Affymetrix U133 microarray. The expression levels of the long variant of COPA are shown in purple, while the expression levels of NCSTN are in yellow. The expression levels in different samples 1-5 are from different microarray experiments.

FIG. 25a is a sequence alignment showing a 3'-3' overlap between IL-24 (GenBank Accession No. NM_006850) and TOSO (GenBank Accession No. NM_005449).

FIG. 25b shows automatic annotation for IL-24 and TOSO obtained using the annotation platform described in Example 10 of the Examples section, which follows.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of methods of identifying naturally occurring antisense transcripts, which can be used in kits and methods for quantifying gene

19

expression levels. Specifically, the antisense molecules and related oligonucleotides generated according to information derived therefrom of the present invention can be used to detect, quantify, or specifically regulate antisense and respective sense transcripts thereby enabling detection and treatment of a wide range of disorders.

5   The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following
10   description or illustrated in the drawings described in the Examples section. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

15   ***Terminology***

As used herein, the term "oligonucleotide" refers to a single stranded or double stranded oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics thereof. This term includes oligonucleotides composed of naturally-occurring bases, sugars and covalent internucleoside linkages (e.g.,
20   backbone) as well as oligonucleotides having non-naturally-occurring portions, which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases.

25   The term "antisense" refers to a complementary strand of an mRNA transcript e.g., antisense RNA.

The phrase "naturally occurring antisense transcripts" refers to RNA transcripts encoded from an antisense strand of the DNA. These endogenous transcript exhibit at least partial complementarity to mRNA transcripts transcribed
30   from the sense strand of a DNA, also termed sense transcripts. *cis*-encoded naturally occurring antisense transcripts are transcribed from the same locus as the sense

20

transcripts. *trans*-encoded antisense transcripts are transcribed from a different locus than the respective sense transcripts.

The phrase "antisense strand" or "anticoding strand" refers to a strand of DNA, which serves as a template for mRNA transcription and as such is complementary to the mRNA transcript formed.

The phrase "sense strand" or "coding strand" refers to the strand of DNA, which is identical to the mRNA transcript formed.

The phrase "complementary DNA" (cDNA) refers to the double stranded or single stranded DNA molecule, which is synthesized from a messenger RNA template.

The phrase "sense oriented polynucleotides" refers to polynucleotide sequences of a complementary or genomic DNA. Such polynucleotide sequences can be from exon regions, in which case they can encode mRNAs or portions thereof, or from intron regions, in which case they typically do not encode mRNA or portions thereof.

The term "contig" refers to a series of overlapping sequences with sufficient identity to create a longer contiguous sequence.

The term "cluster" refers to a plurality of contigs all derived, with a high degree of probability, from a single gene. Clusters are generally formed based upon a specified degree of homology and overlap (e.g., a stringency). The different contigs in a cluster do not typically represent the entire sequence of the gene, rather the gene may comprise one or more unknown intervening sequences between the defined contigs.

The phrase "open reading frame" (ORF) refers to a nucleotide sequence, which could potentially be translated into a polypeptide. Such a stretch of sequence is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons. For the purposes of this application, an ORF may be any part of a coding sequence, with or without start and/or stop codons. For an ORF to be considered as a good candidate for coding for a bona fide cellular protein, a minimum size requirement is often set, for example, a stretch of DNA that would code for a protein of 50 amino acids or more. An ORF is not usually considered an equivalent to a gene

21

or locus until a phenotype is associated with a mutation in the ORF, an mRNA transcript for a gene product generated from the ORF's DNA has been detected, and/or the ORF's protein product has been identified.

The term "annotation" refers to a functional or structural description of a sequence, which may include identifying attributes such as locus name, poly(A)/poly(T) tail and/or signal, key words, Medline references and orientation cloning data.

Naturally occurring antisense molecules can play a role in sense transcription stability and function (e.g. translation). To date, most, if not all of the information relating to naturally occurring antisense transcripts was obtained by either low efficiency computational approaches (described hereinabove) or by approaches utilizing RNase protection assays, northern blot analysis, strand-specific RT PCR, subtractive hybridization, differential plaque hybridization, affinity chromatography, electrospray mass spectrometry and the like. These methods, though highly reliable, are extremely laborious, time consuming and are directed at individual target transcripts. As such, current approaches for uncovering antisense transcripts can be used to detect a negligible portion of the number of naturally occurring antisense molecules thought to exist.

As described hereinunder and in the Examples section, which follows, the present invention provides a novel approach for systematically identifying naturally occurring antisense molecules.

Aside from large scale applicability, the present method can be used to identify naturally occurring antisense molecules even in cases where the antisense transcriptional unit is localized to an intron of an expressed gene or to a different locus than the complementary sense encoding gene (e.g., trans-encoded antisense), or in cases where the antisense molecule lacks an open reading frame or appreciable complementarity to known sense molecules. Antisense transcripts uncovered according to the teachings of the present invention can be used for detecting and accurately quantifying respective sense counterparts as well as for sensibly designing artificial antisense molecules suitable for down-regulation of sense counterparts.

Thus, according to one aspect of the present invention there is provided a method of identifying putative naturally occurring antisense transcripts.

22

The method according to this aspect of the present invention is effected by the following steps.

First, sense-oriented polynucleotide sequences of a first database are computationally aligned with expressed polynucleotide sequences of a second database.

Following computational alignment, expressed polynucleotide sequences are analyzed according to one or more criteria for their ability to hybridize or form a duplex or partial complementation with the sense-oriented polynucleotide sequences (further detailed hereinbelow and in the Examples section which follows).

Expressed polynucleotide sequences which are capable of forming a duplex with sense oriented sequences are considered as putative naturally occurring antisense molecules and as such can be stored in a database which can be generated by a suitable computing platform.

Final confirmation of computationally obtained putative naturally occurring antisense molecules can be effected either computationally or preferably by using suitable laboratorial methodologies, based on nucleotide hybridization including RNase protection assay, subtractive hybridization, differential plaque hybridization, affinity chromatography, electrospray mass spectrometry, northern analysis, RT-PCR and the like (for further details see the Examples section).

Information derived from the sequence, sense position and other structure characteristics of the naturally occurring antisense transcripts identified according to the teachings of the present invention can be used to quantify respective sense transcripts of interest or to generate corresponding artificial antisense polynucleotides, which can be packed in diagnostic or therapeutic kits and implemented in various therapeutic and diagnostic methods.

Expressed polynucleotide sequences used as a potential source for identifying naturally occurring antisense transcripts according to this aspect of the present invention are preferably libraries of expressed messenger RNA [i.e., expressed sequence tags (EST), cDNA clones, contigs, pre-mRNA, etc.] obtained from tissue or cell-line preparations which can include genomic and/or cDNA sequence.

Expressed polynucleotide sequences, according to this aspect of the present invention can be retrieved from pre-existing publicly available databases (i.e.,

23

GenBank database maintained by the National Center for Biotechnology Information (NCBI), part of the National Library of Medicine, and the TIGR database maintained by The Institute for Genomic Research) or private databases (i.e., the LifeSeq.™ and PathoSeq.™ databases available from Incyte Pharmaceuticals, Inc. of Palo Alto, CA).

Alternatively, the sequence database of the expressed polynucleotide sequences utilized by the present invention can be generated from sequence libraries (e.g., cDNA libraries, EST libraries, mRNA libraries and others). cDNA libraries are suitable sources for expressed sequence information.

Generating a sequence database in such a case is typically effected by tissue or cell sample preparation, RNA isolation, cDNA library construction and sequencing.

It will be appreciated that such cDNA libraries can be constructed from RNA isolated from whole organisms, tissues, tissue sections, or cell populations. Libraries can also be constructed from tissue reflecting a particular pathological or physiological state. Of particular interest are libraries constructed from sources associated with certain disease states, including malignant, neoplastic, hyperplastic tissues and the like.

Once raw sequence data is obtained, sequences are selected and preferably annotated before stored in a database. Selection proceeds according to one or more sequence criterion, which will be further detailed hereinunder. The editing, annotation and selection process is divided into two stages of processing. One stage comprises removal of repetitive, redundant or non-informative and contaminant sequences. The second stage involves selection of suitable candidates of putative naturally occurring antisense sequences.

The following section describes the different selection criteria which can be used for sequence filtering.

Vector contamination - "chops" vector elements and linker motifs used for the process of cloning from desired expressed nucleotide sequences. This selection can be effected by screening manually updated databases of sequences included in commonly used expression or cloning vectors.

Contaminating sequences - includes sequences which are derived from an undesired source. Such sequences can be recognized by their nucleotide distribution and/or by homology searches such as alignment searches using any sequence

alignment algorithm such as BLAST (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST) or the Smith-Waterman algorithm. Other contaminating sequences may include sequences exhibiting high occurrence of di-nucleotide distribution mostly related to sequencing artifacts and ribosomal RNA sequences.

Repetitive elements and low complexity sequences - eliminates or masks expressed sequences comprising known repetitive elements (ALU, L1 etc.) and low complexity sequences (i.e., a di- or tri-nucleotide repeat). Such elimination is preferably effected by comparison with database of known repetitive elements. It will be appreciated that this type of selection is mostly species specific. Masking of low complexity sequences can be effected by substituting an N (i.e., an inert character) for the actual nucleotide (i.e., G, A, T, or C). Masking of low complexity sequences facilitates further computational analysis and maintains the spacing of the molecule.

Sequence length - preferred expressed sequences are of a length between 20–2000, preferably 20-1000, more preferably 20-500, most preferably 20-300 base pairs.

Sequence annotation – expressed sequences retrieved from external databases, i.e., GenBank, oftentimes include an annotation which indicates direction of the sequencing of the insert clone (i.e., 5' or 3' direction). Sequence annotation, though "noisy" by nature due to multiple entries from various sources; artifacts taking place during directional cloning and incidence of palindromic eight-cutter restriction sites located at the end of the sequence, can serve as an important tool for deducing strand identity using dedicated computer software which are further discussed hereinunder

Intron splice site consensus sequence intron splice site sharing– intron sequences nearly always begin with a di-nucleotide sequence of GT ("splice donor") and end with an AG ("splice acceptor") preceded by a pyrimidine-rich tract. This consensus sequence is part of the signal for splicing. Intron splice site consensus sequence on the complementary strand (e.g., antisense strand) begins with CT and ends with AC. Thus, combined with genomic data, expressed sequences having a GT...AG can be considered as sense-oriented sequences, while a CT...AC pattern is considered as an antisense oriented sequence. This selection criterion is very stringent since only negligible portions of introns have a CT...AC pattern. Sequences

25

that share a similar splicing pattern, as deduced by alignment to genomic data, may be considered as having the same sense orientation, also termed herein as "intron sharing". It will be appreciated by one skilled in the art that using these selection criteria requires a careful and accurate alignment of expressed sequences to genomic sequence.

Poly(A) tails and Poly(T) heads – most eukaryotic mRNA molecules contain a poly-adenylation [poly(A)] tail at their 3' end. This poly(A) tail is not encoded by DNA. Therefore an expressed sequence which has a poly(A) tail can be considered as sense oriented. Similarly, poly(T) heads, which are not encoded from a genomic sequence indicate that a sequence is of the opposite direction, namely antisense oriented. Notably, genomically encoded Poly(A) tails and poly(T) heads provide no information as to the sequence orientation.

Poly(A) signal – some mature mRNA transcripts contain internal AAUAAA sequence. This internal sequence is part of an endonuclease cleavage signal. Following cleavage by the endonuclease, a poly(A) polymerase adds about 250 A residues to the 3' end of the transcript. Hence, expressed sequences containing a poly(A) signal can be considered as sense oriented.

Rare restriction site used for cloning– for example, eight cutter endonucleases which cleave 8-mer palindromic sequences and are characterized by a low frequency of cutting often used in genome mapping and EST library preparations (e.g., NotI. Commercially available from Promega: www.promega.com). Therefore, when a cluster of overlapping expressed sequences is characterized by a portion of sequences starting with a digestion site and another portion ending with the same, these sequences may be considered as encoded from the same strand. However, any endonuclease capable of digesting a palindromic sequence (i.e., XhoI, SalI, PacI etc.) may also affect distorted sequence clustering, therefore strand orientation is preferably effected using other parameters as well.

Sequence overlap - sequences that completely overlap are considered to have the same strand orientation.

The above-described parameters are used individually or in combination to analyze the expressed polynucleotide sequences so as to select anti-sense oriented sequences.

26

Selection can be effected on the basis of a single criterion or several criteria considered individually or in combination.

In cases where several criteria are examined, a scoring system e.g., a scoring matrix, is preferably used.

Since in some cases identifying an intron splicing consensus site may be more important than both sequence annotation and NotI alignment, while in others, detection of poly(A) tails and poly(T) heads might be the most significant criterion, the use of a scoring matrix in which each criterion is weighted enables one to select qualified antisense transcripts.

Such a scoring matrix can list the various expressed polynucleotide sequences across the X-axis of the matrix while each criterion can be listed on the Y-axis of the matrix. Criteria include both a predetermined range of values from which a single value is selected from each sequence, and a weight. Each sequence is scored at each criterion according to its value and the weight of the criterion.

When using such a scoring matrix the scores of each criterion of a specific sequence are summed and the results are analyzed.

Expressed sequences which exhibit a total score greater than a particular stringency threshold are grouped as members of either a sense-oriented sequence set or antisense-oriented sequence set; the higher the score the more stringent the criteria of grouping.

It will be appreciated that the above-described analysis can take place prior to computational alignment to sense oriented sequences, i.e., during the process of editing the expressed sequence database, which is, described hereinabove. Alternatively, selection can take place following computational alignment, thus further facilitating identification of proper duplex formation between the sense oriented polynucleotide sequences and expressed polynucleotide sequences.

Genomic DNA or a portion thereof is preferably used as sense-oriented sequence data according to this aspect of the present invention. It is conceivable that the present invention can determine sense orientation and antisense orientation of a database of expressed sequences simply by computationally aligning the sequences of the expressed database onto the genome, and finding whether two complementary expressed sequences hybridize to the genome (e.g., virtually generate a double

27

stranded portion thereof). Such two overlapping sequences constitute sense and naturally occurring antisense transcripts.

Utilizing genomic DNA as a sense oriented template is preferred for the following reasons: (i) identifying trans-encoded antisense transcripts; (ii) analyzing intron splice consensus site and intron sharing; (iii) omitting genomically encoded poly(A) and poly(T) sequences; and (iv) analyzing sequences encompassing eight-cutter restriction sites.

Computational alignment of expressed polynucleotide sequences to the sense-oriented polynucleotide sequences (e.g., genomic sense sequences) can be effected using any commercially available alignment software, including sequence alignment tools utilizing algorithm such as BLAST (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST) or Smith-Waterman.

Assembly software is preferably used according to this aspect of the present invention. Such software is of high value when complete genomic information is unavailable or when handling large amounts of expressed sequence data. A number of commonly used computer software fragment read assemblers capable of forming clusters of expressed sequences are now available. These packages include but are not limited to, The TIGR Assembler [Sutton G. et al. (1995) Genome Science and Technology 1:9-19], GAP [Bonfield JK. et al. (1995) Nucleic Acids Res. 23:4992-4999], CAP2 [Huang X. et al. (1996) Genomics 33:21-31], The Genome Construction Manager [Laurence CB. Et al. (1994) Genomics 23:192-201], Bio Image Sequence Assembly Manager, SeqMan [Swindell SR. and Plasterer JN. (1997) Methods Mol. Biol. 70:75-89], LEADS and GenCarta (Compugen Ltd. Israel).

Computer assembly and alignment programs can be modified to incorporate sequence criteria for determining sense or antisense orientation of expressed nucleotide sequences, as described hereinabove. Thereby, avoiding deliberate inversion of sequences during the assembly process, while ignoring the natural orientation of the sequences (i.e., sense or antisense orientation). Figure 1 illustrates results of expressed sequence assembly against genomic data and final distinction between sense oriented transcripts and antisense oriented transcripts of a single gene.

Following a proper alignment of expressed sequences to sense oriented polynucleotide sequences, duplexes are identified. The term "duplex" is used herein

28

to indicate that a sequence identified according to this aspect of the present invention is complementary to a sense-oriented polynucleotide sequence. Complementation may be to a portion of the sense sequence, i.e., a region thereof, or alternatively, to two or more non-contiguous regions, which may be separated by one or more
5   nucleotides on the sense strand.

The formation of sense-antisense duplexes does not require 100 % complementation nor does it require participation of the entire sense/antisense transcript sequence. The sense or antisense transcripts can have a secondary structure (e.g., stem and loop) generated by intra-sequence hybridization which can prevent
10   specific sequence regions in the sense or antisense transcripts from participating in duplex formation. Thus, the antisense of the sequence identified, according to this aspect of the present invention can be complementary to its sense counterparts in several regions, which are not necessarily close to each other when the sense transcript is in linear form.

15   Although any length of sequence overlap can generate a duplex, overlaps of at least about 5, preferably about 20, more preferably about 30, even more preferably about 40 bp are considered more indicative of true sense-antisense duplex formation.

The method of uncovering putative antisense transcripts of the present invention is preferably carried out using a dedicated computational system.

20   Thus, according to another aspect of the present invention and as illustrated in Figure 2, there is provided a system for generating a database of putative naturally occurring antisense sequences which system is referred to hereinunder as system 10.

System 10 includes at least one central processing unit (CPU) 12, which executes a software application designed and configured for aligning sense oriented
25   polynucleotide sequences with expressed polynucleotide sequences and identifying expressed polynucleotide sequences which are capable of forming a duplex with the sense oriented polynucleotide sequences, thereby recognizing putative naturally occurring antisense transcripts. System 10 may also include a user input interface 14 [e.g., a keyboard and/or a cursor control device (e.g., a joy stick)] for inputting
30   database or database related information, and a user output interface 16 (e.g., a monitor) for providing database information to a user.

System **10** may also include random access memory **24**, ROM memory **26**, a modem **28** and a graphic processing unit (GPU) **30**.

System **10** preferably stores sequence information of the putative antisense transcripts identified thereby on an internal and/or external storage device **20** such as a magnetic, optico-magnetic or optical disk as a database of putative antisense transcript sequences. Such a database further includes information pertaining to database generation (e.g., source library), parameters used for selecting polynucleotide sequences, putative uses of the stored sequences, and various other annotations (as described below) and references which relate to the stored sequences or respective sense transcripts.

The hardware elements of system **10** may be tied together by a common bus or several interlinked buses for transporting data between the various elements. Examples of system **10** include but are not limited to, a personal computer, a work station, a mainframe and the like.

System **10** of the present invention may be used by a user to query the stored database of sequences, to retrieve nucleotide sequences stored therein or to generate polynucleotide sequences from user inputted sequences.

The methods of the present invention can be effected by any software application executable by system **10**. The software application can be stored in random access memory **24**, or internal and/or external data storage device **20** of system **10**.

The database generated and stored by system **10** can be accessed by an on-site user of system **10**, or by a remote user communicating with system **10**, through for example, a terminal or thin client.

The latter configuration is best exemplified by the client-server system **50** which is shown in Figure 3. System **50** is configured to perform similar functions to those performed by system **10**. In system **50**, communication between a remote client **18** (e.g., computer, PDA, cell phone etc) and processing unit **12** of a local server or computer is typically effected via a communication network **32**. Communication network **32** can be any private or public communication network including, but not limited to, a standard or cellular telephony network, a computer

network such as the Internet or intranet, a satellite network or any combination thereof.

As illustrated in Figure 3, communication network 32 can include one or more communication servers 22 (one shown in Figure 3) which serve for communicating data pertaining to the sequence of interest between remote client 18 and processing unit 12. Thus, a request for data or processed data is communicated from remote client 18 to processing unit 12 through communication network 32 and processing unit 12 sends back a reply which includes data or processed data to remote client 18. Such a system configuration is advantageous since it enables users of system 50 to store and share gathered information and to collectively analyze gathered information.

Such remote configuration can be implemented over a local area network (LAN) or a wide area network (WAN) using standard communication protocols.

It will be appreciated that existing computer networks such as the Internet can provide the infrastructure and technology necessary for supporting data communication between any number of users 18 and processors 12.

Novel polynucleotide sequences uncovered using the above-described methodology can be used in various clinical applications (e.g., therapeutic and diagnostic) as is further described hereinbelow.

A polynucleotide sequence of the present invention refers to a single or double stranded nucleic acid sequence which is isolated and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequence (e.g., a combination of the above.

As used herein the phrase "complementary polynucleotide sequence" refers to a sequence, which results from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such a sequence can be subsequently amplified *in vivo* or *in vitro* using a DNA dependent DNA polymerase.

As used herein the phrase "genomic polynucleotide sequence" refers to a sequence derived (isolated) from a chromosome and thus it represents a contiguous portion of a chromosome.

As used herein the phrase "composite polynucleotide sequence" refers to a sequence, which is composed of genomic and cDNA sequences. A composite sequence can include some exonal sequences required to encode the polypeptide of the present invention, as well as some intronic sequences interposing therebetween. The

5   intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

Thus, the present invention encompasses nucleic acid sequences described hereinabove; fragments thereof, sequences hybridizable therewith, sequences

10   homologous thereto (e.g., at least about 50 %, at least about 55 %, at least about 60%, at least about 65 %, at least about 70 %, at least about 75 %, at least about 80 %, at least about 85 %, at least about 95 % or more say 100 % homologous to the nucleic acid sequences set forth in the file seqs_125 or seqs_133 of the enclosed CD-ROM1), sequences encoding similar polypeptides with different codon usage, altered

15   sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

In cases where the polynucleotide sequences of the present invention encode previously unidentified polypeptides, the present invention also encompasses novel

20   polypeptides or portions thereof, which are encoded by the isolated polynucleotide and respective nucleic acid fragments thereof described hereinabove.

Thus, the present invention also encompasses polypeptides encoded by the polynucleotide sequences of the present invention. The present invention also encompasses homologues of these polypeptides, which can optionally be at least

25   about 50 %, at least about 55 %, at least about 60%, at least about 65 %, at least about 70 %, at least about 75 %, at least about 80 %, at least about 85 %, at least about 95 % or more say 100 % homologous to the amino acid sequences set forth in the file pep_seqs_133 of the enclosed CD-ROM3 Finally, the present invention also encompasses fragments of the above described polypeptides and polypeptides having

30   mutations, such as deletions, insertions or substitutions of one or more amino acids, either naturally occurring or man induced, either randomly or in a targeted fashion.

32

Thus, data extracted from the database of naturally occurring antisense transcripts of the present invention is of high value for designing oligonucleotides suitable for transcript detection and quantification and for sensibly designing artificial antisense oligonucleotides for down-regulation and elimination of a transcript of interest or changing the balance between sense and complementary antisense transcripts. The possibility of up-regulating a transcript of interest using naturally occurring antisense based-oligonucleotides generated according to the teachings of the present invention is also realized. In addition, data extracted from the database of naturally occurring antisense transcripts may also be used for assessing endogenous double stranded-RNA also termed interfering RNA, which may distort gene-expression due to either RNA-degradation, DNA-methylation, polycomb mediated suppression etc. (for details see the Background section hereinabove).

Antisense technology is based upon the pairing of an artificially designed antisense oligonucleotide, with a target nucleic acid. The use of antisense technology requires a complementarity of the antisense nucleotide sequence to a target zone of an mRNA target sequence that will effect inhibition of gene expression [reviewed in Stein CA. and Cohen JS. (1988) Cancer Res. 48:2659-68]. Based on empiric experience it was shown that the success of antisense technology relies on: (i) cellular uptake; (ii) stability of artificial antisense molecules under physiological conditions (i.e., cellular pH, endonucleases etc.); (iii) complementation between the oligonucleotide and a single stranded target sequence (i.e., tertiary structure of target RNA will not form a good target); (iv) binding specificity of antisense oligonucleotide so as not to compete with other RNA binders (e.g. proteins) to thereby maintain an effective antisense concentration.

Various attempts to employ antisense technology while considering the above discussed limitations included using large amounts of oligonucleotides to overcome cellular uptake and environmental barriers and chemically modified antisense nucleotide compositions, for obtaining higher level of cellular stability. However, even in case where uptake difficulties are traversed, the step of target identification (i.e., RNA-target sequence region) continues to be the major bottleneck for successful implementation of antisense technology.

U.S. Pat. No: 6,183,966 discloses a method and an apparatus for ranking nucleic acid sequences based on stability of nucleic acid oligomer sequence binding interactions to select sequence zones for antisense targeting. This method however systematic, relies on thermodynamic analyses combined with numerous predictions which cannot be considered empirically accurate and reliable.

Thus according to another aspect of the present invention there is provided a method of designing artificial antisense transcripts.

The method according to this aspect of the present invention is effected by the following steps.

First, structural and/or functional parameters pertaining to naturally occurring antisense transcripts are extracted/deduced from a database such as the one described hereinabove. These parameters may be generally deduced from all sequences stored in the database, or extracted from specific antisense sequences or preferably groups of antisense sequences.

Second, artificial antisense molecules of interest are designed according to the extracted parameters.

Such parameters may be divided into three groups, topographical parameters, functional parameters and structural parameters.

Topographical parameters - (i) position of sequence overlap on the sense transcript (i.e., coding region, 5'UTR, 3'UTR); (ii) position of the sequence overlap on the antisense transcript (end overlap, middle overlap, full overlap). (iii) length of overall sequence overlap; (iv) continuity or discontinuity of sequence overlap.

Structural parameters - pertains to both sense and antisense transcripts (i) tertiary structure (i.e., hairpin, helix, stem and loop, pseudoknot, and the like); (ii) single stranded versus double stranded regions; (iii) GC content; (iv) tandem Gs; (v) adenosine/inosine content; (vi) thermodynamic stability of tertiary structures; (vii) duplex melting point; (viii) methylations and other RNA modifications; (ix) RNA-protein interactions ; and (x) transcript length.

Functional parameters - (i) alternative splicing; (ii) tissue expression; (iii) pathology specific expression; (iv) antisense promoters; (v) intron content; (vi) open reading frame in antisense transcript.

34

These parameters can be used individually or in combination, in which case, each parameter is preferably weighted according to its importance. Due to the multifactorial design of artificial antisense transcripts according to this aspect of the present invention, employing a scoring system (described hereinabove) is preferably used to simplify and increase the accuracy of the process.

Synthetic antisense oligonucleotides designed according to the teachings of the present invention can be generated according to any oligonucleotide synthesis method known in the art such as enzymatic synthesis or solid phase synthesis. Equipment and reagents for executing solid-phase synthesis are commercially available from, for example, Applied Biosystems. Any other means for such synthesis may also be employed; the actual synthesis of the oligonucleotides is well within the capabilities of one skilled in the art.

Oligonucleotides used according to this aspect of the present invention are those having a length selected from a range of from about 10 to about 200 bases, preferably from about 15 to about 150 bases, more preferably from about 20 to about 100 bases, most preferably from about 20 to about 50 bases.

The oligonucleotides of the present invention may comprise heterocylic nucleosides consisting of purines and the pyrimidines bases, bonded in a 3' to 5' phosphodiester linkage.

Preferably used oligonucleotides are those modified in either backbone, internucleoside linkages or bases, as is broadly described hereinunder. Such modifications can oftentimes facilitate oligonucleotide uptake and resistance to intracellular conditions.

Specific examples of preferred oligonucleotides useful according to this aspect of the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone, as disclosed in U.S. Pat. NOs: ,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466, 677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkyl phosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms can also be used.

Alternatively, modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH2 component parts, as disclosed in U.S. Pat. Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623, 070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439.

Other oligonucleotides which can be used according to the present invention, are those modified in both sugar and the internucleoside linkage, i.e., the backbone, of the nucleotide units are replaced with novel groups. The base units are maintained for complementation with the appropriate polynucleotide target. An example for such an oligonucleotide mimetic, includes peptide nucleic acid (PNA). A PNA oligonucleotide refers to an oligonucleotide where the sugar-backbone is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The

bases are retained and are bound directly or indirectly to aza nitrogen atoms of the
amide portion of the backbone. United States patents that teach the preparation of
PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,082; 5,714,331;
and 5,719,262, each of which is herein incorporated by reference. Other backbone

5   modifications, which can be used in the present invention are disclosed in U.S. Pat.
No:6,303,374.

Oligonucleotides of the present invention may also include base modifications
or substitutions. As used herein, "unmodified" or "natural" bases include the purine
bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine

10 (C) and uracil (U). Modified bases include but are not limited to other synthetic and
natural bases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine,
xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of
adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-
thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl

15 uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-
thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted
adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-
substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-
azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-

20 deazaguanine and 3-deazaadenine. Further bases include those disclosed in U.S. Pat.
No: 3,687,808, those disclosed in The Concise Encyclopedia Of Polymer Science And
Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those
disclosed by Englisch et al., Angewandte Chemie, International Edition, 1991, 30,
613, and those disclosed by Sanghvi, Y. S., Chapter 15, Antisense Research and

25 Applications, pages 289-302, Crooke, S. T. and Lebleu, B. , ed., CRC Press, 1993.
Such bases are particularly useful for increasing the binding affinity of the oligomeric
compounds of the invention.   These include 5-substituted pyrimidines, 6-
azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-
aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine

30 substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 °C.
[Sanghvi YS et al. (1993) Antisense Research and Applications, CRC Press, Boca

Raton 276-278] and are presently preferred base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more moieties or conjugates, which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-S-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety, as disclosed in U.S. Pat. No: 6,303,374.

It is not necessary for all positions in a given oligonucleotide molecule to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide.

The present invention also includes antisense molecules, which are chimeric molecules. "Chimeric" antisense molecules", are oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one nucleotide. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target polynucleotide. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. An example for such include RNase H, which is a cellular endonuclease which cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Cleavage of the RNA target can be routinely detected by gel electrophoresis and, if necessary, associated nucleic acid hybridization techniques known in the art.

38

Chimeric antisense molecules of the present invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, as described above. Representative U.S. patents that teach the preparation of such hybrid structures include, but are not limited to, U.S. Pat. Nos. 5,013,830; 5,149,797; 5,220,007; 5,256,775; 5,366,878; 5,403,711; 5,491,133; 5,565,350; 5,623,065; 5,652,355; 5,652,356; and 5,700,922, each of which is herein fully incorporated by reference.

Finally, chimeric oligonucleotides of the present invention can comprise a ribozyme sequence. Ribozymes are being increasingly used for the sequence-specific inhibition of gene expression by the cleavage of mRNAs. Several ribozyme sequences can be fused to the oligonucleotides of the present invention. These sequences include but are not limited ANGIOZYME specifically inhibiting formation of the VEGF-R (Vascular Endothelial Growth Factor receptor), a key component in the angiogenesis pathway, and HEPTAZYME, a ribozyme designed to selectively destroy Hepatitis C Virus (HCV) RNA, (Ribozyme Pharmaceuticals, Incorporated - WEB home page).

Naturally occurring antisense sequences uncovered using the teachings of the present invention can be annotated using information available from a number of publicly available sources of gene annotations well known to those of skill in the art. Examples include, but are not limited to Locus Link and RefSeq: GO annotations, Gencarta (described in Example 10 of the Examples section), GeneCards, GeneLynx, TIGR and the like.

Annotative information obtained using the Gencarta (Compugen, Tel-Aviv, Israel) database is set forth in the file "annotations_133" of the enclosed CD-ROM3. The relation between the co-regulated nucleotide sequences of the present invention and corresponding polypeptide sequences is described in Trans2Pep file in the enclosed CD-ROM3.

Elucidating protein function, pattern of expression, therapeutic and diagnostic roles, allows for the design of highly specific and effective clinical tools, for a wide range of diseases as described in the Examples section which follows.

For example, gene products (nucleic acid and/or protein products), which exhibit tumor specific expression (i.e., tumor associated antigens, TAAs) can be

utilized for in-vitro generation of antibodies and/or for in-vivo immunization/cancer vaccination, essentially eliciting an immune response against such gene products and cells expressing same (see e.g., U.S. Pat. No. 4,235,877, Vaccine preparation is generally described in, for example, M. F. Powell and M. J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995); Other references describing adjuvants, delivery vehicles and immunization in general include Rolland, Crit. Rev. Therap. Drug Carrier Systems 15:143-198, 1998; Fisher-Hoch et al., Proc. Natl. Acad. Sci. USA 86:317-321, 1989; Flexner et al., Ann. N.Y Acad. Sci. 569:86-103, 1989; Flexner et al., Vaccine 8:17-21, 1990; U.S. Pat. Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Pat. No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, Biotechniques 6:616-627, 1988; Rosenfeld et al., Science 252:431-434, 1991; Kolls et al., Proc. Natl. Acad. Sci. USA 91:215-219, 1994; Kass-Eisler et al., Proc. Natl. Acad. Sci. USA 90:11498-11502, 1993; Guzman et al., Circulation 88:2838-2848, 1993; and Guzman et al., Cir. Res. 73:1202-1207, 1993; Ulmer et al., Science 259:1745-1749, 1993; Cohen, Science 259:1691-1692, 1993; U.S. Pat. Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094; U.S. Pat. Nos. 6,008,200 and 5,856,462; Zitvogel et al., Nature Med. 4:594-600, 19980.

Tumor-specific gene products of the present invention, in particular membrane bound, can be utilized as targeting molecules for binding therapeutic toxins, antibodies and small molecules, to thereby specifically target the tumor cell. Alternatively, neoplastic properties of tumor specific gene products (nucleic acid and/or protein products) of the present invention, may be beneficially used in the promotion of wound healing and neovascularization in ischemic conditions and diabetes.

Secreted variants of known autoantigens associated with a specific autoimmune syndrome, such as for example, those listed in Table 11, below, can be used to treat such syndromes. Typically, autoimmune disorders are characterized by a number of different autoimmune manifestations (e.g., multiple endocrine syndromes). For these reasons secreted variants may be used to treat any combination of autoimmune phenomena of a disease as detailed in Table 11 below. The therapeutic effect of these variants may be a result of (i) competing with autoantigens for binding with autoantibodies; (ii) antigen-specific immunotherapy, essentially suggesting that

40

systemic administration of a protein antigen can inhibit the subsequent generation of the immune response to the same antigen (proven in mice models for Myasthenia Gravis and type I Diabetes).

Biomolecular sequences, which are over-expressed in a pathology can be used as diagnostic markers, such as for cancer. Variants of autoantigens may also be used for diagnosis. The diagnosis of many autoimmune disorders is based on identification of autoantibodies specific to autoantigens known to be associated with an autoimmune condition. Most of the diagnostic techniques are based on having a recombinant form of the autoantigen and using it to look for serum autoantibodies. It is possible that currently considered autoantigens are not "true" autoantigens but rather variants thereof. For example, TPO is a known autoantigen for thyroid autoimmunity. It has been shown that its variant TPOzanelli also takes part in the autoimmune process and can bind the same antibodies as TPO [Biochemistry. 2001 Feb 27;40(8):2572-9.]. Antibodies formed against the true autoantigen may bind to other variants of the same gene due to sequence overlap but with reduced affinity. Novel splice variant of the genes in Table 11 may be revealed as true autoantigens, therefore their use for detection of autoantibodies is expected to result in a more sensitive and specific test.

Additionally, variants of known drug targets can be used in cases where the known drug has major side effects, the therapeutic efficacy of the known drug is moderate, the drug failed clinical trials due to one of the above. A drug which is specific to a new protein variant of the target or to the target only (without affecting the novel variant) is likely to have less side effects as compared with the original drug, higher efficacy and may treat different indications than the original drug.

For example, COX3, which is a variant of COX1, is known to bind COX inhibitors with a different affinity than COX1. This molecule is also associated with different physiological processes than COX1. Therefore, a compound specific to COX1 or compounds specific to COX3 would have lower side effects (by not affecting the other variant), treat different indications and treat successfully bigger populations.

41

Polynucleotide sequence data (i.e., mRNAs and/or naturally occurring antisense transcripts thereof) obtained according to the teachings of the present invention may also be used for modulating the expression of a gene of interest.

For example, when a product of a gene of interest is an oncogene, it may be desirable to upregulate an expression of a naturally occurring antisense transcript thereof, which is in turn capable of downregulating expression of the gene product (i.e., the oncogene), to thereby inhibit the onset or progression of a malignant disease. See also Example 11 of the Examples section which follows.

Alternatively, when a gene product of interest is a tumor suppressor gene it may be desirable to downregulate expression of a naturally occurring antisense molecule thereof to thereby elevate expression of the tumor suppressor gene and inhibit the onset or progression of a malignant disease.

Upregulation or downregulation of the naturally occurring antisense transcript of the gene product of interest, as described above, may be accompanied by upregulation or down regulation of the gene product itself as necessary (e.g., upregulation of a tumor suppressor gene and down regulation of an oncogene).

Agents for upregulating and down regulating gene products (i.e., mRNA or proteins) are summarized infra.

Upregulating expression of a naturally occurring antisense transcript of interest may be effected via the administration of at least one of the exogenous polynucleotide sequences of the present invention, ligated into a nucleic acid expression construct designed for expression of coding sequences in eukaryotic cells (e.g., mammalian cells). Accordingly, the exogenous polynucleotide sequence may be a DNA or RNA sequence encoding the naturally occurring antisense transcript of interest.

For therapeutic applications, the nucleic acid construct can be administered to an individual in need therefore by employing any suitable mode of administration described hereinbelow (i.e., in-vivo gene therapy). Alternatively, the nucleic acid construct can be introduced into an isolated cells, of for example, a cell culture, using an appropriate gene delivery vehicle/method (transfection, transduction, homologous recombination, etc.). The genetically modified cells thus generated can then be expanded in culture and returned to the individual (i.e., ex-vivo gene therapy).

42

To enable cellular expression of the polynucleotides of the present invention, the nucleic acid construct of the present invention further includes at least one cis acting regulatory element. As used herein, the phrase "cis acting regulatory element" refers to a polynucleotide sequence, preferably a promoter, which binds a trans acting

5   regulator and regulates the transcription of a coding sequence located downstream thereto.

Any suitable promoter sequence can be used by the nucleic acid construct of the present invention.

Preferably, the promoter utilized by the nucleic acid construct of the present

10   invention is active in the specific cell population transformed. Examples of cell type-specific and/or tissue-specific promoters include promoters such as albumin that is liver specific [Pinkert et al., (1987) Genes Dev. 1:268-277], lymphoid specific promoters [Calame et al., (1988) Adv. Immunol. 43:235-275]; in particular promoters of T-cell receptors [Winoto et al., (1989) EMBO J. 8:729-733] and immunoglobulins;

15   [Banerji et al. (1983) Cell 33729-740], neuron-specific promoters such as the neurofilament promoter [Byrne et al. (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477], pancreas-specific promoters [Edlunch et al. (1985) Science 230:912-916] or mammary gland-specific promoters such as the milk whey promoter (U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). The nucleic acid

20   construct of the present invention can further include an enhancer, which can be adjacent or distant to the promoter sequence and can function in up regulating the transcription therefrom.

The nucleic acid construct of the present invention preferably also includes an appropriate selectable marker and/or an origin of replication. Preferably, the nucleic

25   acid construct utilized is a shuttle vector, which can propagate both in *E. coli* (wherein the construct comprises an appropriate selectable marker and origin of replication) and be compatible for propagation in cells, or integration in a gene and a tissue of choice. The construct according to the present invention can be, for example, a plasmid, a bacmid, a phagemid, a cosmid, a phage, a virus or an artificial

30   chromosome.

Examples of suitable constructs include, but are not limited to, pcDNA3, pcDNA3.1 (+/-), pGL3, PzeoSV2 (+/-), pDisplay, pEF/myc/cyto, pCMV/myc/cyto

43

each of which is commercially available from Invitrogen Co. (www.invitrogen.com). Examples of retroviral vector and packaging systems are those sold by Clontech, San Diego, Calif., including Retro-X vectors pLNCX and pLXSN, which permit cloning into multiple cloning sites and the trasgene is transcribed from CMV promoter.

5   Vectors derived from Mo-MuLV are also included such as pBabe, where the transgene will be transcribed from the 5'LTR promoter.

Currently preferred in vivo nucleic acid transfer techniques include transfection with viral or non-viral constructs, such as adenovirus, lentivirus, Herpes simplex I virus, or adeno-associated virus (AAV) and lipid-based systems. Useful

10   lipids for lipid-mediated transfer of the gene are, for example, DOTMA, DOPE, and DC-Chol [Tonkinson et al., Cancer Investigation, 14(1): 54-65 (1996)]. The most preferred constructs for use in gene therapy are viruses, most preferably adenoviruses, AAV, lentiviruses, or retroviruses. A viral construct such as a retroviral construct includes at least one transcriptional promoter/enhancer or locus-defining element(s),

15   or other elements that control gene expression by other means such as alternate splicing, nuclear RNA export, or post-translational modification of messenger. Such vector constructs also include a packaging signal, long terminal repeats (LTRs) or portions thereof, and positive and negative strand primer binding sites appropriate to the virus used, unless it is already present in the viral construct. In addition, such a

20   construct typically includes a signal sequence for secretion of the peptide from a host cell in which it is placed. Preferably the signal sequence for this purpose is a mammalian signal sequence or the signal sequence of the polypeptide variants of the present invention. Optionally, the construct may also include a signal that directs polyadenylation, as well as one or more restriction sites and a translation termination

25   sequence. By way of example, such constructs will typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second-strand DNA synthesis, and a 3' LTR or a portion thereof. Other vectors can be used that are non-viral, such as cationic lipids, polylysine, and dendrimers.

When the naturally occurring antisense transcript and/or gene product of

30   interest are splice isoforms it is possible to promote expression of a specific splice variant by using antisense oligonucleotides, which are directed at splice sites of interest, thereby altering the splicing pattern of the gene. This approach has been

44

successfully used for shifting the balance of expression of the two isoforms of Bcl-x [Taylor (1999) Nat. Biotechnol. 17:1097-1100; and Mercatante (2001) J. Biol. Chem. 276:16411-16417]; IL-5R [Karras (2000) Mol. Pharmacol. 58:380-387]; and c-myc [Giles (1999) Antisense Acid Drug Dev. 9:213-220].

5   For example, interleukin 5 and its receptor play a critical role as regulators of hematopoiesis and as mediators in some inflammatory diseases such as allergy and asthma. Two alternatively spliced isoforms are generated from the IL-5R gene, which include (i.e., long form) or exclude (i.e., short form) exon 9. The long form encodes an intact membrane-bound receptor, while the shorter form encodes a secreted soluble 10   non-functional receptor. Using 2'-O-MOE-oligonucleotides specific to regions of exon 9, Karras and co-workers (supra) were able to significantly decrease the expression of the wild type receptor and increase the expression of the shorter isoforms. Approaches which can be used to design and synthesize oligonucleotides according to the teachings of the present invention are described hereinbelow and by 15   Sazani and Kole (2003) Progress in Moleclular and Subcellular Biology 31:217-239.

Upregulation at the protein level may be effected by providing the polypeptide gene product (i.e., either of the gene of interest or of the naturally occurring antisense transcript). It will be appreciated that since the bioavailability of large polypeptides is relatively small due to high degradation rate and low penetration rate, administration 20   of polypeptides is preferably confined to small peptide fragments (e.g., about 100 amino acids).

Oligonucleotides for downregulating gene expression are described hereinabove.

Antibodies can also be used to down regulate activity of a gene product (i.e., 25   protein) of interest. Methods of producing polyclonal and monoclonal antibodies as well as fragments thereof are well known in the art. See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1988, incorporated herein by reference. Other methods for the generation of antibody fragments include for example, U.S. Pat. Nos. 4,036,945   and 4,331,647, and 30   references contained therein; Porter, R. R. Biochem. J. 73: 119-126 (1959); Proc. Nat'l Acad. Sci. USA 69:2659-62 (1972); Whitlow and Filpula, Methods 2: 97-105 (1991); Bird et al., Science 242:423-426 (1988); Pack et al., Bio/Technology 11:1271-

45

77 (1993); and U.S. Pat. No. 4,946,778. References for methods of humanizing non-human antibodies include Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988); U.S. Pat. No. 4,816,567; Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et

5   al., J. Mol. Biol., 222:581 (1991). The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies [Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

The present invention also envisages enhancing a therapeutic efficacy of a

10   therapeutic agent.

As used herein a "therapeutic agent" refers to a polynucleotide, a polypeptide or a small molecule, which targets a gene product of interest to induce a therapeutic effect in a subject.

Therapeutic agents (drugs), which are used to treat diseases or disorders

15   function through one of several mechanisms. Typically, therapeutic agents target one or more endogenous molecules to exert a negative or positive effect thereupon. The following section describes several therapeutic modalities in which use of naturally occurring antisense transcripts in combination with the therapeutic agent is advantageous.

20   In cases where the gene product targeted is suppressed or eliminated by the therapeutic agent [e.g., a soluble receptor (i.e., therapeutic agent) which is capable of downregulating an activity of an oncogenic receptor (i.e., target)], the effect of the therapeutic agent can be enhanced by providing to a subject in need thereof a naturally occurring antisense transcript of the gene encoding the expression product

25   targeted by the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent. Such combination therapy can be utilized to suppress the gene target at both the translational and functional level thus substantially enhancing therapeutic efficacy. Methods of administration are further described hereinbelow.

Alternatively, the subject may be provided with a naturally occurring antisense

30   transcript of a gene which encodes an expression product capable of inhibiting activity or expression of the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent. For example, any of the sequences shown in Table

46

9 below, which are of proteins that are involved in drug-drug interactions, may be suitable for selection in this regard. A non-limiting example of such a sequence has the name (gene symbol) "ABCB9". If such a protein was shown to be involved in an interaction which would reduce the efficacy of a therapeutic agent, then the naturally

5   occurring antisense transcript of the corresponding gene could optionally be provided as an adjunct therapy, by being provided to the subject prior to, concomitant with, or following provision of the therapeutic agent.

Still alternatively, the subject may be provided with an inhibitor [e.g., oligonucleotide, small molecule, antibody) of a naturally occurring antisense

10   transcript of a gene which encodes an expression product capable of upregulating activity or expression of the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent. For example, see Example 12 of the Examples section which follows. In this case the naturally occurring antisense transcript performs a dual function, by both downregulating activity of the therapeutic agent at

15   the transcript level and inhibiting activity at the protein level.

Still alternatively, in cases where a therapeutic agent functions by activating an endogenous molecule such as a cell receptor (e.g., VEGF-R); it may be beneficial to upregulate expression or activity of a naturally occurring antisense transcript of an endogenous non-functional form of this receptor (e.g., soluble VEGF-R). Since the

20   non-functional form of the receptor may be capable of competing for the therapeutic agent and thus masking or eliminating its effect on the cell bound (wt) form of the receptor, provision of a naturally occurring antisense transcript of the non-functional form of the target gene can either increase the beneficial effect of the therapeutic agent and/or reduce the amount of therapeutic agent used in treatment.

25   The oligonucleotides and polynucleotides of the present invention can be used for therapeutic purposes. For example, oligonucleotides of the present invention can be used to treat a variety of diseases or pathological conditions associated with an abnormal expression (i.e., up-regulation or down-regulation) of at least one mRNA molecule of interest, including but not limited to diabetes, autoimmune diseases,

30   Parkinson, Alzheimer' disease, HIV, malaria, cholera, influenza, rabies, diphtheria, breast cancer, colon cancer, cervical cancer, melanoma, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, lymphomas, leukemias and the like and any other

diseases (see Example 8 of the Examples section) which are associated with aberrant expression of multiple mRNAs (i.e., sense and/or antisense) or with unregulated formation of endogenous double stranded RNA complexes.

Furthermore, accumulated knowledge strongly correlates human diseases with mutations, over-expression and function of the protein building blocks (i.e., protein kinases, phosphatsases) and their effectors and regulators, which constitute numerous intracellular signaling pathways. For instance, inactivation of both copies of ZAP-70 or Jak-3 causes severe combined immunodeficiency and mutation of the X-linked BTK gene results in agammaglobulinemia. Many genetic disorders are also associated with mutations for example, in protein-serine kinases (PSKs) and phosphatases. The Coffin-Lowry syndrome results from inactivation of the X-linked Rsk2 gene, and myotonic dystrophy is due to decreased levels of expression of the myotonic dystrophy PSK. In addition, over-expression of ErbB2 receptor tyrosine kinase is implicated in breast and ovarian carcinoma [reviewed by Hunter T. (2000) Cell 100:113-127].

Given the importance of activated kinases in a variety of disorders such as cancer, it would be anticipated that phosphatases regulation would be found as tumor suppressor genes and as promising drug targets. So far this has not proved to be the case. Furthermore, a number of diseases are associated with insufficient expression of signaling molecules, including non-insulin-dependent diabetes and peripheral neuropathies.

Thus, it is conceivable that identification of naturally occurring antisense transcripts of signaling molecules participating in specified signaling pathways may serve as promising tools for both identification and particularly treatment of a variety of disorders at any gene expression level (i.e., RNA, DNA or protein).

The term "treating" refers to alleviating or diminishing a symptom associated with the disease or the condition and/or eliminating the cause of the disease or condition. Preferably, treating cures, e.g., substantially eliminates, and/or substantially decreases, the symptoms and/or cause of the disease or condition.

The treatment method according to the teachings of the present invention includes administering to an individual a therapeutically effective amount of the oligonucleotides, polynucleotides or other agents of the present invention. Preferred

48

individual subjects according to the present invention are mammals such as canines, felines, ovines, porcines, equines, bovines, humans and the like.

A therapeutically effective amount implies an amount of agent effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the individual being treated

The agent of the method of the present invention can be administered to an individual *per se*, or as part of a pharmaceutical composition where it is mixed with a pharmaceutically acceptable carrier.

As used herein a "pharmaceutical composition" refers to a composition of one or more of the agents described hereinabove, or physiologically acceptable salts or prodrugs thereof, with other chemical components. The purpose of a pharmaceutical composition is to facilitate administration of a compound to an organism.

The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including vaginal and rectal delivery), pulmonary, e.g., by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration. Oligonucleotides with at least one 2'-O-methoxyethyl modification are believed to be particularly useful for oral administration.

Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Coated condoms, gloves and the like may also be useful.

Compositions and formulations for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. Thickeners, flavoring agents, diluents, emulsifiers, dispersing aids or binders may be desirable.

49

Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

5   Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, and liposome-containing formulations. These compositions may be generated from a variety of components that include, but are not limited to, preformed liquids, self-emulsifying solids and self-emulsifying semisolids.

The pharmaceutical formulations of the present invention, which may
10   conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the formulations are prepared by uniformly and intimately bringing into association the active ingredients with liquid
15   carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, liquid syrups, soft gels, suppositories, and enemas. The compositions of the present
20   invention may also be formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

In one embodiment of the present invention the pharmaceutical compositions
25   may be formulated and used as foams. Pharmaceutical foams include formulations such as, but not limited to, emulsions, microemulsions, creams, jellies and liposomes. While basically similar in nature these formulations vary in the components and the consistency of the final product. The preparation of such compositions and formulations is generally known to those skilled in the pharmaceutical and
30   formulation arts and may be applied to the formulation of the compositions of the present invention.

50

The pharmaceutical compositions of the present invention may employ various penetration enhancers to effect the efficient delivery of nucleic acids, particularly oligonucleotides, to the skin of animals.

Penetration enhancers may be classified as belonging to one of five broad categories, i.e., surfactants, fatty acids, bile salts, chelating agents, and non-chelating non-surfactants [Lee et al., Critical Reviews in Therapeutic Drug Carrier Systems (1991) 92] as disclosed in U.S. Pat. No: 6,300,132, 6,271,030, 6,277,633, 6,284,538, 6,287,860, 6,294,382, 6,277,640 and 6,258,601 each of which is herein fully incorporated by reference.

Other substances that enhance uptake of oligonucleotides at the cellular level may also be added to the pharmaceutical compositions of the present invention. For example, cationic lipids, such as lipofectin [U.S. Pat. No. 5,705,188], cationic glycerol derivatives, and polycationic molecules, such as polylysine [PCT Application WO 97/30731], are also known to enhance the cellular uptake of oligonucleotides.

Other reagents may be utilized to enhance the penetration of the administered nucleic acids, including glycols such as ethylene glycol and propylene glycol, pyrrols such as 2-pyrrol, azones, and terpenes such as limonene and menthone.

Certain pharmaceutical compositions of the present invention may also incorporate carrier compounds. As used herein, "carrier compound" or "carrier" can refer to a nucleic acid, or analog thereof, which is inert (i.e., does not possess biological activity per se) but is recognized as a nucleic acid by in vivo processes that reduce the bioavailability of a nucleic acid having biological activity by, for example, degrading the biologically active nucleic acid or promoting its removal from circulation. The co-administration of a nucleic acid and a carrier compound, typically with an excess of the latter substance, can result in a substantial reduction of the amount of nucleic acid recovered in the liver, kidney or other extracirculatory reservoirs, presumably due to competition between the carrier compound and the nucleic acid for a common receptor. For example, the recovery of a partially phosphorothioate oligonucleotide in hepatic tissue can be reduced when it is coadministered with polyinosinic acid, dextran sulfate, polycytidic acid or 4-acetamido-4' isothiocyano-stilbene-2,2'-disulfonic acid [Miyao et al., Antisense Res.

Dev., (1995) 5:115-121; Takakura et al., Antisense & Nucl. Acid Drug Dev. (1996) 6:177-183].

In contrast to a carrier compound, an "excipient" is a pharmaceutically acceptable solvent, suspending agent or any other pharmacologically inert vehicle for delivering one or more nucleic acids to an animal. The excipient may be liquid or solid and is selected, with the planned manner of administration in mind, so as to provide for the desired bulk, consistency, etc., when combined with a nucleic acid and the other components of a given pharmaceutical composition. Typical excipients include, but are not limited to, binding agents (e.g., pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose, etc.); fillers (e.g., lactose and other sugars, microcrystalline cellulose, pectin, gelatin, calcium sulfate, ethyl cellulose, polyacrylates or calcium hydrogen phosphate, etc.); lubricants (e.g., magnesium stearate, talc, silica, colloidal silicon dioxide, stearic acid, metallic stearates, hydrogenated vegetable oils, corn starch, polyethylene glycols, sodium benzoate, sodium acetate, etc.); disintegrants (e.g., starch, sodium starch glycolate, etc.); and wetting agents (e.g., sodium lauryl sulphate, etc.).

Pharmaceutically acceptable organic or inorganic excipient suitable for non-parenteral administration which do not deleteriously react with nucleic acids can also be used to formulate the compositions of the present invention. Suitable pharmaceutically acceptable carriers include, but are not limited to, water, salt solutions, alcohols, polyethylene glycols, gelatin, lactose, amylose, magnesium stearate, talc, silicic acid, viscous paraffin, hydroxymethylcellulose, polyvinylpyrrolidone and the like.

Formulations for topical administration of nucleic acids may include sterile and non-sterile aqueous solutions, non-aqueous solutions in common solvents such as alcohols, or solutions of the nucleic acids in liquid or solid oil bases. The solutions may also contain buffers, diluents and other suitable additives. Pharmaceutically acceptable organic or inorganic excipients suitable for non-parenteral administration, which do not deleteriously react with nucleic acids can be used.

Suitable pharmaceutically acceptable excipients include, but are not limited to, water, salt solutions, alcohol, polyethylene glycols, gelatin, lactose, amylose,

52

magnesium stearate, talc, silicic acid, viscous paraffin, hydroxymethylcellulose, polyvinylpyrrolidone and the like.

The compositions of the present invention may additionally contain other adjunct components conventionally found in pharmaceutical compositions, at their art-established usage levels. Thus, for example, the compositions may contain additional, compatible, pharmaceutically-active materials such as, for example, antipruritics, astringents, local anesthetics or anti-inflammatory agents, or may contain additional materials useful in physically formulating various dosage forms of the compositions of the present invention, such as dyes, flavoring agents, preservatives, antioxidants, opacifiers, thickening agents and stabilizers. However, such materials, when added, should not unduly interfere with the biological activities of the components of the compositions of the present invention. The formulations can be sterilized and, if desired, mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, colorings, flavorings and/or aromatic substances and the like which do not deleteriously interact with the nucleic acid(s) of the formulation. Aqueous suspensions may contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

The formulation of therapeutic compositions and their subsequent administration is believed to be within the skill of those in the art. Dosing is dependent on severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on EC50 found to be effective in in vitro and in vivo animal models. Persons of ordinary skill in the art can easily estimate dosing and repetition rates based on measured residence times and concentrations of the oligonucleotide in bodily fluids or tissues. Following successful treatment, it may be desirable to have the patient

53

undergo maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is administered in maintenance doses.

As mentioned hereinabove, biomolecular sequences of the present invention can also be used in diagnostic applications. Present-day mRNA-based diagnostic assays utilize oligonucleotide probes which are complementary to one or more regions of the mRNA to be qualified or quantified. Such probes are designed while considering interspecies sequence variation, sequence length, GC content etc. However design of such prior art probes (i.e., riboprobes or deoxyriboprobes) does not take into consideration the presence of antisense transcripts which can effect probe binding efficiency. Discounting antisense presence can lead to inaccurate qualification/quantification and thus diagnosis, and thus lead to erroneous treatment protocols.

The present invention provides an mRNA-detection/quantification assay, which is devoid of this limitation.

Thus, according to an additional aspect of the present invention there is provided a method of quantifying at least one mRNA of interest in a biological sample.

As used herein, the phrase "biological sample" refers to any sample derived from biological tissues or fluids, including blood (serum or plasma), sputum, pleural effusions, urine, biopsy specimens, isolated cells and/or cell membrane preparation. Methods of obtaining tissue biopsies and body fluids from mammals are well known in the art.

The method of this aspect of the present invention is effected by contacting mRNA from a cell type or within a cell with one or more oligonucleotides that hybridizes efficiently with a sequence region of an mRNA transcript which is not complementary with a naturally occurring antisense transcript.

In addition to the limitation described above, prior art diagnostic/detection assays also fail to consider the effect of antisense transcription on the protein expression levels of a gene of interest. It stands to reason that presence of antisense transcripts in a biological sample can substantially reduce the resultant protein levels translated from a complementary sense transcript. Consistently, diseases which are associated with endogenous dsRNA complexes, are also very difficult to detect and

moreover to treat, due to insufficient sequence data pertaining to duplex forming transcripts.

Thus, for accurate quantification of gene expression, both the sense and antisense levels must be quantified and/or their respective expression ratio must be determined.

By contacting a biological sample with one or more pairs of oligonucleotides, where one oligonucleotide is capable of hybridizing with the mRNA of interest and the second oligonucleotide is capable of hybridizing with a naturally occurring antisense transcript which is complementary with the mRNA of interest such accurate quantification can be effected.

Contacting the oligonucleotides of the present invention with the biological sample is effected by stringent, moderate or mild hybridization (as used in any polynucleotide hybridization assay such as northern blot, dot blot, RNase protection assay, RT-PCR and the like). Wherein stringent hybridization is effected by a hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 1 - 1.5 °C below the Tm, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the Tm; moderate hybridization is effected by a hybridization solution of 6 x SSC and 0.1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 2 - 2.5 °C below the Tm, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the Tm, final wash solution of 6 x SSC, and final wash at 22 °C; whereas mild hybridization is effected by a hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 mg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 37 °C, final wash solution of 6 x SSC and final wash at 22 °C.

The oligonucleotides of the present invention can be attached to a solid substrate, which may consist of a particulate solid phase such as nylon filters, glass slides or silicon chips [Schena et al. (1995) Science 270:467-470].

55

In a particular embodiment, oligonucleotides of the present invention can be attached to a solid substrate, which is designed as a microarray. Microarrays are known in the art and consist of a surface to which probes that correspond in sequence to gene products (e.g., cDNAs, mRNAs, cRNAs, polypeptides, and fragments
5   thereof), can be specifically hybridized or bound at a known position (regiospecificity).

Several methods for attaching the oligonucleotides to a microarray are known in the art including but not limited to glass-printing, described generally by Schena et al., 1995, Science 270:467-47, photolithographic techniques [Fodor et al. (1991)
10   Science 251:767-773], inkjet printing, masking and the like.

In general, quantifying hybridization complexes is well known in the art and may be achieved by any one of several approaches. These approaches are generally based on the detection of a label or marker, such as any radioactive, fluorescent, biological or enzymatic tags or labels of standard use in the art. A label can be
15   applied on either the oligonucleotide probes or nucleic acids derived from the biological sample.

The following illustrates a number of labeling methods suitable for use in the present invention. For example, oligonucleotides of the present invention can be labeled subsequent to synthesis, by incorporating biotinylated dNTPs or rNTP, or
20   some similar means (e.g., photo-cross-linking a psoralen derivative of biotin to RNAs), followed by addition of labeled streptavidin (e.g., phycoerythrin-conjugated streptavidin) or the equivalent.   Alternatively, when fluorescently-labeled oligonucleotide probes are used, fluorescein, lissamine, phycoerythrin, rhodamine (Perkin Elmer Cetus), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX (Amersham) and
25   others [e.g., Kricka et al. (1992), Academic Press San Diego, Calif] can be attached to the oligonucleotides. It will be appreciated that pairs of fluorophores are chosen when distinction between two emission spectra of two oligonucleotides is desired or optionally, a label other than a fluorescent label is used. For example, a radioactive label, or a pair of radioactive labels with distinct emission spectra, can be used [Zhao
30   et al. (1995) Gene 156:207]. However, because of scattering of radioactive particles, and the consequent requirement for widely spaced binding sites, the use of fluorophores rather than radioisotopes is more preferred.

56

The intensity of signal produced in any of the detection methods described hereinabove may be analyzed manually or using a software application and hardware suited for such purposes.

In general, mRNA quantification is preferably effected alongside a calibration
5   curve so as to enable accurate mRNA determination. Furthermore, quantifying transcript(s) originating from a biological sample is preferably effected by comparison to a normal sample, which sample is characterized by normal expression pattern of the examined transcript(s).

It will be appreciated that the detection method described above can also be
10 used for quantifying at least one naturally occurring antisense transcript in a biological sample. In such a case, the oligonucleotide used for quantification is designed to hybridize with a sequence region of naturally occurring antisense transcript of interest, which is not complementary with a naturally occurring mRNA transcript.

15   The diagnostic assays described hereinabove can be used to accurately distinguish between absence, presence and excess expression of any transcripts of interest (e.g., sense, antisense), and to monitor their level during therapeutic intervention. These methods are also capable of diagnosing diseases associated with an improper balance or ratio between sense and antisense expression and diseases
20 associated with endogenous dsRNA.

Further description of oligonucleotide-pair arrays is provided in Example 9 of the Examples section which follows.

As discussed hereinabove oligonucleotides of the present invention can be also used for therapeutic purposes, such as treating diseases or conditions associated
25 with aberrant expression levels of one or more sense and/or antisense transcripts and conditions, which are associated with endogenous dsRNA such as unregulated formation of double-strand RNA (i.e., up/down-regulation).

The oligonucleotides generated according to the teachings of the present invention can be included in a diagnostic or therapeutic kit. For example,
30 oligonucleotides sets pertaining to specific disease related transcripts can be packaged in a one or more containers with appropriate buffers and preservatives along with

57

suitable instructions for use and used for diagnosis or for directing therapeutic treatment.

Preferably, the containers include a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers may be formed from a variety of materials such as glass or plastic.

In addition, other additives such as stabilizers, buffers, blockers and the like may also be added.

Apart from clinical applications, the biomolecular sequences of the present invention can find other commercial uses such as in the food, agricultural, electromechanical, optical and cosmetic industries [www.physics.unc.edu/~rsuper/XYZweb/XYZchipbiomotors.rs1.doc; www.bio.org/er/industrial.asp]. For example, newly uncovered gene products, which can disintegrate connective tissues, can be used as potent anti scarring agents for cosmetic purposes. Non-limiting examples of such gene products include the matrix metalloproteinase family of proteins (MMP), which are a group of proteases having varying specificities for ECM components as substrates, non-limiting examples of which have the gene symbols "CLG" and "CGL4B" in the attached files. These proteins are involved in ECM break-down as part of the wound healing process, for example for cell migration. The activity of these proteins is also modulated by specific tissue inhibitors of MMPs (TIMP) and other factors in the microenvironment in and around the wound area. Therefore, one possible optionally application for the present invention would be the selection of appropriate antisense oligonucleotides for either one or more MMPs and /or for factors related to TIMPs, in order to modulate wound healing activities (and/or as previously noted, for treatment of arthritis).

As another optional treatment, production of collagen may be optionally modulated through the use of appropriate antisense oligonucleotides. Collagen is an important connective tissue element, but is also involved in pathological conditions such as fibrosis and the formation of adhesions between tissues of different organs, a condition which may occur for example after surgery. Therefore, modulation of collagen production, for example to reduce collagen production, may optionally be performed according to the present invention.

58

Other applications include, but are not limited to, the making of gels, emulsions, foams and various specific products, including photographic films, tissue replacers and adhesives, food and animal feed, detergents, textiles, paper and pulp, and chemicals manufacturing (commodity and fine, e.g., bioplastics).

5   Research applications include, for example, differential cloning, detection of rearrangements in DNA sequences as disclosed in U.S. Pat. No: 5,994,320, drug discovery and the like.

As used herein the term "about" refers to ± 10 %.

10

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove

15   and as claimed in the claims section below finds experimental support in the following examples.

EXAMPLES

Reference is now made to the following examples, which together with the

20   above descriptions, illustrate the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et

25   al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory

30   Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III

59

Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak et al., "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

In-vitro expression substantiation of computationally retrieved naturally occurring antisense transcripts

In-vitro expression assays were conducted in order to validate the existence of naturally occurring antisense sequences identified according to the teachings of the present invention.

Table 1 below lists polynucleotide sequence pairs that were selected for the *in-vitro* expression validation assays described in examples 1-7.

Table 1

Name of sense antisense pair	Sense transcript	Sense Length (nt)	Antisense transcript	Anti- sense Length (nt)	Overlap length (nt)	Start of overlap sense transcript	Start of overlap anti- sense

60

53BP1_76P	53BP1	10394	76P	6837	3046	5463	2018
	(SEQ ID NO: 15)		(SEQ ID NO: 16)				
CIDEB_BLTR2 (1)	CIDEB1	2289	BLTR2	6530	2254	17	1
	(SEQ ID NO: 19)		(SEQ ID NO: 21)				
CIDEB_BLTR2 (2)	CIDEB2	1511	BLTR2	6530	1410	1	1
	(SEQ ID NO: 20)						
APAF1_EB1	aAPAF1	7042	EB1a	1752	141	6889	1612
	(SEQ ID NO: 24)		(SEQ ID NO: 25)				
AChR_MINK2	AchR	2457	MINK2	4863	236	2175	4853
	(SEQ ID NO: 29)		(SEQ ID NO: 30)				
M-AchR_Anti-AChR	M-AchR	1590	M-Anti-AchR	2227	672	934	506
	(SEQ ID NO: 35)		(SEQ ID NO: 36)				
CyclinE2_Anti-CyclinE2	CyclinE2	2714	Anti-CyclinE2	5773	1855	565	2006
	(SEQ ID NO: 33)		(SEQ ID NO: 34)				

Sequence alignments of overlapping regions of each sense-antisense pair were performed using the BLAST sequence alignment algorithm (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST using the default parameters) and are exhibited in Figure 5a-g.

A microarray-based analysis was conducted, as well, in order to validate the existence of naturally occurring, antisense sequences identified according to the teachings of the present invention. The results are described in Example 9.

Materials and Experimental Methods

RNA probes generation and northern analysis

RNA probes for northern analysis were generated by PCR amplification of a desired DNA fragment and cloning into Zero Blunt TOPO (Invitrogen Corp.) or pSPT18/19 vectors (Roche Ltd.). Alternatively PCR products were ligated into T7 RNA polymerase promoter-containing adaptors using the Lignscribe kit (Ambion Europe Ltd.). Corresponding RNA transcripts were synthesized using T7 RNA polymerase (Roche Ltd.) and labeled with 32P-UTP according to manufacturer's instructions. RNA probes were purified on Mini Quick Spin RNA columns.

Commercial membranes containing Poly(A)-RNA from various human tissues (2 μg RNA per lane) were obtained from Origene (OriGene Technologies Inc.) and Ambion (Ambion Inc.).

Alternatively, 2 μg of poly(A)-RNA prepared from various human cell-lines were electrophoretically separated on 1 % agarose gel, and electrotransferred to Nytran SuperCharge membrane (Schleicher & Schuell) and subjected to fixing by UV radiation. Membranes were stained with methylene blue to ensure quantitative RNA transfer. Membranes were then prehybridized in a hybridization solution (UltraHyb solution Ambion Europe Ltd.) for 30 minutes at 68 °C in a rotating hybridization tube.

Hybridization solution was then supplemented with 106 cpm of labeled RNA probe per each ml of hybridization solution. Blots were hybridized for 16 hours at 68 °C in a rotating hybridization tube. Membranes were then washed twice with 2 x SSC, 0.1 % sodium dodecyl sulfate (SDS) and twice with 0.1 % SDS at 68 °C. RNA transcripts signals were detected using a phosphoimager (Molecular Dynamics, Sunnyvale CA).

Microarray

Oligonucleotide design - oligonucleotide design tools (1) were applied to each pair of sense/antisense genes in order to select two complementary 60-mer oligonucleotides from the region where the two genes overlap. The design criteria included the following: low cross-homology (up to 75%) to other expressed sequences in the human transcriptome; a continuous hit of no more than 17 bp to the sequence of another gene; balanced GC content (30-70%) without significant windows of local imbalance; no more than 2 palindromes with a length of 6 bp; a hit of no more than 15 bp to a repeat, vector or low-complexity region; and no long stretches of identical nucleotides.

Microarray preparation - 60-mer oligonucleotides were synthesized by Sigma-Genosys (The Woodlands, TX), resuspended at 40 μM in 3X SSC, and spotted in quadruplicates on poly-L-lysine coated glass slides as detailed in the online protocol of the National Human Genome Research Institute (http://www.nhgri.nih.gov/DIR/Microarray/Protocols.pdf). To avoid local differences in the hybridization conditions, the probes selected from the overlapping regions of each sense/antisense pair were spotted in the same block, next to each other.

Human cell lines - The following cell lines utilized were purchased from ATCC (Manassas, VA): MCF7 (breast adenocarcinoma, Cat. No. HTB-22,), HeLa

62

(cervical adenocarcinoma, Cat. No. CCL-2) HEK-293 (embryonal kidney cells, Cat.
No. CRL-1573), Jurkat (acute T-cell leukemia, Cat. No. TIB-152), K-562 (chronic
myelogenous leukemia, Cat. No. CCL-243), HepG2 (liver carcinoma, Cat. No. HB-
8065), T24 (urinary bladder carcinoma, Cat. No. HTB-4), SK-N-DZ (neuroblastoma,
5   Cat. No. CRL-2149), NK-92 (non-Hodgkin's lymphoma, Cat. No. CRL-2407), MG-63
(osteosarcoma, Cat. No. CRL-1427), DU 145 (prostatic carcinoma, Cat. No. HTB-81),
G-361 (melanoma, Cat. No. CRL-1424), PANC-1 (pancreatic carcinoma, Cat. No.
CRL-1469), ES-2 (ovary clear cell carcinoma, Cat. No. CRL-1978), Y79
(retinoblastoma, Cat. No. HTB-18), HT-29 (colorectal adenocarcinoma, Cat. No.
10   HTB-38), H1299 (large cell lung carcinoma, Cat. No. CRL-5803), SNU1 (gastric
carcinoma, Cat. No. CRL-5971), NL564 (EBV-transformed human lymphoblasts) and
MCF10 (benign tumor breast cells).

      RNA purification - Total RNA was extracted from the above mentioned
human cell lines using TriReagent (Molecular Research Center, Cincinnati, OH).
15   Poly(A)+ mRNA was purified using two cycles of the Dynabeads mRNA Purification
Kit (Dynal Biotech ASA, Oslo, Norway), as per manufacturer instructions. The
removal of traces of ribosomal RNA was confirmed by agarose gel electrophoresis.
Poly(A)+ mRNAs from human testis, placenta, lung and brain tissue were purchased
from BioChain Institute, Inc. (Hayward, CA). mRNAs of all cell lines described
20   above were combined in equal quantities to obtain the reference 'mRNA pool'.

      Preparation of labeled cDNA - For each hybridization, labeled cDNA was
synthesized by reverse transcription of 0.5 μg of mRNA, in the presence of 100 pmol
of random 9-mers, 1μg of oligo(dT)20, 1X RT buffer, 10 mM DTT, 3 nmol of Cy5- or
Cy3-conjugated dUTP, 0.5 mM of dATP, dGTP and dCTP, and 0.2 mM dTTP, in a
25   final volume of 40 μl (Amersham). The reaction mixture was incubated for 5 minutes
at 65 °C and cooled to 42 °C.   600 Units of reverse transcriptase (Superscript II,
Invitrogen, Carlsbad, CA) and 40 U of Rnase inhibitor (RNasin Promega, Madison,
WI) were added and the reaction was incubated for 30 minutes at 42 °C. An
additional 200 U of Superscript II were added and the reaction was incubated for
30   another 15 minutes. Remaining RNA was degraded by the addition of 200 mM
NaOH and 50 mM EDTA, at 65 °C for 10 minutes. The mixture was neutralized by
adding half a volume of 1M Tris-HCl pH 7.5. Hybridizations were performed in

63

duplicate using fluorescent reversal of Cy3- and Cy5-labeled cDNA from test cell mRNAs and pooled mRNAs. Pairs of Cy5/Cy3-labeled cDNA samples were combined, and subsequently purified and concentrated to a final volume of 5-7 µl using a Microcon-30 (Millipore) concentrator.

5   *Hybridization and washing conditions* - Microarray slides were prehybridized with 40 µl of 5X SSC, 0.1 % SDS and 1 % BSA for 30 min at 42 °C, washed for 2 minutes with double distilled water, then rinsed with isopropanol, and spun dried at 500 g for 3 minutes. Prior to hybridization, the labeled probe was combined with 10 µg of Cot-1 DNA, 10 µg poly(dA)80, and 4 µg yeast tRNA, in a final volume of 15

10   µl. The mixture was denatured at 100 °C for 3 minutes and placed on ice. Formamide (final concentration 16 %), SSC (to 5X concentration) and 0.1 % SDS were added to a final volume of 30 µl. The mixture was placed on the array under a glass cover slip in a tightly sealed hybridization chamber, and immersed in a water bath at 42 °C, for 16 hours. Microarray slides were then washed for 4 minutes with 2X SSC, 0.1 % SDS; 4

15   minutes with 1X SSC, 0.01 % SDS; 4 minutes with 0.2X SSC and 15 seconds with 0.05X SSC and spun dry by centrifugation for 3 minutes at 500g.

Image processing - Following hybridization, arrays were scanned using a GenePix 4000B scanner (Axon Instruments, Union City, CA). Scanned array images were manually inspected and areas with visible artifacts or deformities were marked.

20   Images were processed using GenePix Pro 3.0 (www.axon.com) software.

Normalization - The intensity for each spot was calculated as its mean intensity minus the median background around the spot. The signal for each oligo was calculated as the average of intensity values of the four redundant spots of each oligo. Normalization of the oligo signals was performed at several levels as is further

25   described below.

Normalization of blocks was carried out in order to normalize the gradient of intensities within each slide. For each block i, an Ai parameter was calculated as the average of intensities of 56 positive control spots (oligonucleotide probes for the ubiquitously expressed housekeeping genes gapdh, actin, hsp70 and gnb2ll, in

30   various probe concentrations). An average A of all Ai averages was calculated. Based on this, a block normalization factor Bi was calculated for each block, as Bi = A/Ai, and applied to each spot in the block.

64

Normalization between slides was performed to bring all experiments to the same scale. For each experiment, the average of intensities of the 192 negative control spots on the array was set to be the 0 (zero) of the new scale. For a subset of highly signaling oligos, with intensities between the 70th and the 95th percentiles of the

5   oligo signal distribution of the experiment, the average was arbitrarily set to be 500 in the new scale. The intensity of each oligo signal was accordingly converted to this new scale, to obtain the normalized signal. A ratio between the normalized cell-line signal and the normalized pool signal was calculated for each oligo in each experiment. To avoid misleading ratios coming from signals that were too low, the

10 ratio Rji for oligo j in experiment i was calculated as: $Rji = \max [100, \text{cell-line-signalji}]/\max [100, \text{pool-signalji}]$.

To normalize between red/green intensities in reciprocal experiments, the ratio Rjk for oligo j in cell-line k was calculated as the average of calculated ratios Rji between the two reciprocal experiments of the cell-line k. In cases where only one of

15 the two reciprocal experiments showed an elevated or decreased ratio, while in the other the ratio was 1.0, the average Rjk was converted to 1.0.

The actual pool signal for each oligo was calculated to be the average of the normalized oligo signals in the pool channel of all experiments. A virtual pool signal was calculated as the average of the normalized oligo signals in the cell-line channel

20 of all experiments. The virtual pool signals were found to be very close to the actual pool signals, indicating consistency in the analysis.

Threshold determination - To determine an expression threshold above, in which a normalized signal would be considered a 'positive' signal indicating expression, the distribution of all 16,512 normalized negative control signals and the

25 standard deviation (neg-std-dev) were calculated. The neg-std-dev obtained was 38. An oligo j was considered 'present' in a cell-line k if $Rjk \times \text{actual-pool-signalj} \geq 4 \times \text{neg-std-dev}$.

EXAMPLE 1

30 *Identification of 53BP1 and 76P RNA transcripts in a variety of human tissues and cell-lines*

Background:

65

The tumor suppressor p53 binding protein 1 (SEQ ID NO: 15) is one of the various p53 target proteins. It binds to the DNA-binding domain of p53 and enhances p53-mediated transcriptional activation. 53BP1 is characterized by several structural motifs shared by several proteins involved in DNA repair and/or DNA damage-signaling pathways. 53BP1 becomes hyperphosphorylated and forms discrete nuclear foci in response to DNA damage induced by radiation and chemotherapy. Recent reports suggest that 53BP1 is an ataxia telangiectasia mutated (ATM) substrate that is involved early in the DNA damage-signaling pathways in mammalian cells, attributing a role to 53BP1 in the development of various mammalian pathologies.

Results:

Two 53BP1 RNA sense transcripts with dissimilar 3' UTRs were previously described [Iwabuchi K. et al. (1994) Proc. Natl. Acad. Sci. USA] and are illustrated in Figure 6 (red and green). Leads assembly program modified to uncover novel antisense transcripts was used to uncover three such transcripts for the 53BP1 gene, which transcripts have different 3' UTRs (SEQ ID NO: 16, 37 and 38) and encode the 76p gene product (Genbank accession number NM014444)(illustrated in blue).

To confirm expression of computationally retrieved antisense transcripts, two RNA-probes were generated. Schematic location of the probes used for sense and antisense validation (Riboprobe#1 and Riboprobe#2, respectively SEQ ID NO: 17 and 18, respectively) is illustrated in Figure 6. These RNA probes were used to identify the corresponding full-length transcripts.

As shown in Figure 7, Riboprobe#1 detected two transcripts of approximately 6.3 Kb and 10.5 Kb, corresponding to the sense mRNA. The absolute levels of the short messenger were rather homogeneous in all cell-lines examined. The 10.5 Kb variant exhibited a more heterogenic pattern of cellular distribution, and was mostly expressed in K562, MG-63, 293 HEK and Hela cells. In general, the longer sense transcript which is an alternatively polyadenylated variant was markedly lower expressed in the various cell lines examined.

The same membrane was used to perform northern analysis with Riboprobe#2 in order to validate expression of antisense transcripts of 53BP1. Results are shown in Figure 8. Three variants corresponding to the 76p gene were detected in most of the cell lines: 6.8 Kb, 4.2 Kb and 2.5 Kb. Minor fluctuations of expression were

66

observed and the largest transcript was expressed at significantly higher levels than the smaller transcripts.

A sense strand probe was used to detect expression of the antisense transcripts in a variety of human tissues (Figure 9). The three alternatively polyadenylated variants with different 3' UTRs were expressed in most of the tissues. Total levels of these transcripts varied in the different tissues assayed. For example, highest level of expression for all three transcripts was observed in the brain and testis, while no expression of the 6.8 Kb and 4.2 Kb variants was detected in the spleen. Expression levels of each transcript were summarized in Table 2 below.

Table 2

	Transcript Mol. Weight (Kb)		
Tissue	6.8	4.2	2.5
brain	+++	++++	++++
colon	+	++	+
heart	-	+	++
kidney	++	++	+
Liver	-	-	+
lung	++++	+++	+
muscle	++	+	+
placenta	+	++	++
Small intestine.	++	++	-
spleen	-	-	+
stomach	-	-	+
testis	++	++	++++

Reverse transcription amplification (RT-PCR) analysis was performed in order to substantiate the northern blot results. Primers were synthesized according to the scheme shown in Figure 10 (indicated by arrows). The expected amplification products corresponded completely to the observed amplification reaction products, supporting the existence of the various 53BP1 and 76p transcription variants.

EXAMPLE 2

Identification of mRNA and complementary transcripts of the Cell death inducing DFF45-like effector (CIDE)-B

Background:

67

Cell death inducing DFF45-like effector (CIDE-B) (GenBank Accession numbers AF190901 and AF218586) is a member of a novel family of apoptosis-inducing factors that share homology with the N-terminal region of DFF, the DNA fragmentation factor. Although the molecular mechanism of CIDE-B induced apoptosis in unclear, mitochondrial localization and dimerization, both where shown to be required [Chen Z. et al. (2000) J. Biol. Chem. 275:22619-22622]. Notably, over-expression of CIDE-B in mammalian cells shows strong cell death-inducing activity, suggesting that aberrant expression of this protein may be associated with a number of mammalian pathologies [Inohara N. et al. (1998) EMBO J. 17:2526-2533].

Results:

Two sense transcript of the CIDE-B gene were previously described with different 5' UTRs [Inohara N. et al. (1998) EMBO J. 17:2526-2533 and Lugovskoy AA. et al. (1999) Cell 99:745-755] (SEQ ID NOs: 19 and 20). Computational analysis recovered a potential elongated BLTR2 transcript (SEQ ID NO: 21), showing full complementary to the CIDE-B mRNA transcripts (Figure 11).

Northern blot analysis was done in order to determine the distribution of the CIDE-B sense and antisense transcripts in various cell-lines. A 430 base pairs DNA fragment was selected to generate RNA probes for identification of both sense and antisense transcripts (SEQ ID NOs: 22 and 23, respectively).

Expression of antisense mRNA transcripts was detected in various cell-lines and especially in the mammary gland adenocarcinome cell line-MCF-7 as a predominant 6.5 Kb transcript, although higher forms were also visualized (Figure 12). Low hybridization with a CIDE-B probe was detected (Figure 13).

Conclusion:

BLTR2 was recently identified as a putative seven-transmembrane receptor with a high homology to the Leukotriene B (4) receptor [Tryselius Y. et al. (2000) Biochem. Biophys. Res. Commun. 274:377-82]. Although the mechanism of action of BLTR2 is poorly understood, it is conceivable that BLTR2 mRNA plays a role in the regulation of CIDE-B apoptotic effector and vice versa.

EXAMPLE 3

68

Identification of mRNA and complementary transcripts of the apoptosis inducing factor APAF-1

Background:

A conserved series of events including cellular shrinkage, nuclear

5   condensation, externalization of plasma membrane phosphatidyl serine, and oligonucleosomal DNA fragmentation characterizes apoptotic cell death. Regardless of the circumstance, induction and execution of apoptotic events require activation of caspases, a family of aspartate-specific cysteine proteinases. Caspase activation may be regulated by the mitochondrion and specifically by the apoptosome consisting of

10   an oligomeric complex of apoptotic protease-activating factor-1 (APAF-1), cytochrome C and dATP. The apoptosome recruits and activates caspase-9, which in turn activates the executioner caspases, caspase-3 and -7. The active executioners kill the cell by proteolysis of key cellular substrates [Zou H. et al. (1999) J. Biol. Chem. 274:11549-11556]. Evasion or inactivation of the mitochondrial apoptosis pathway

15   may contribute to oncogenesis by allowing cell proliferation. In this instance, unregulated cell proliferation may occur by inactivation of APAF-1, which has been suggested to occur via genetic loss or inhibition by HSP-70 and HSP-90. Although aberrant expression of APAF-1 was found in a variety of malignancies (including ovarian epithelial cancer), no link was found to accelerated protein degradation.

20   ### Results:

One RNA transcript has been previously described for APAF-1 [Zou H. et al. (1999) J. Biol. Chem. 274:11549-11556] (SEQ ID NO: 10) (SEQ ID NO: 24). Computational search for natural antisense transcripts has revealed two complementary transcripts for APAF-1 messenger RNA (SEQ ID NOs: 25 and 26).

25   These antisense transcripts include an open reading frame encoding the EB-1 gene (GenBank accession numbers AF145204; AF164792). The overlap between the APAF-1 messenger RNA and the longer antisense transcript is of at least 300 nucleotides.

To validate expression of computationally retrieved antisense transcripts for

30   APAF-1, as well as expression of APAF-1 mRNA in the assayed human cell lines, RNA-probes of 366 ribonucleotides were generated (sense and antisense strands, respectively). Schematic location of the probes used for sense and antisense

validation (Riboprobe#1 and Riboprobe#2, SEQ ID NOs: 27 and 28, respectively) is illustrated in Figure 14.

As shown in Figure 15a, the sense RNA probe directed at visualizing the antisense transcripts, identified a clear band of 3 Kb corresponding to the long computationally retrieved antisense transcript as well as other transcripts sizing from 1 Kb to 8 Kb (Figure 15a). Transcripts were essentially found in all cell lines but especially in 293 HEK and LN-Cap lines.

Hybridization with an RNA probe directed at visualizing the mRNA transcript of APAF-1 resulted only in a blurred patterns (Figure 15b). However, a 7 Kb mRNA transcript consistent with APAF-1mRNA was seen in Ln Cap and 293 HEK cell lines.

Conclusion:

A reciprocal pattern of expression was observed for both APAF-1 and EB-1 transcripts, exhibiting an interesting expressional relationship between the sense and antisense transcripts suggesting antisense-mediated expression regulation.

EXAMPLE 4

mRNA expression of muscle nicotinic Acetyl-Choline Receptor ε subunit and its complementary MINK transcript

Background:

The muscle nicotinic Acetylcholine Receptor ε subunit (AChRε) encodes for one of five subunits of a ligand gated ion channel receptor located at the neuromuscular synapse. AChRε is up-regulated in the postnatal period when it replaces □ subunit of the receptor [Witzamann, V. et al., (1987) FEBS Lett. 223, 104-112]. It is also up-regulated in synapse development, specifically by the trophic factor neuregulin [Martinou J. C. (1991) Pro. Natl. Acad. Sci. USA 88, 7669-7673]. In an attempt to decipher AchRε function and mechanism of regulation, computational screen for AChRε K complementary transcript was carried out.

Results:

One mRNA transcript of AChRε gene was previously described [Beeson D. Eur. J. Biochem (1993) 215, 229-238] (SEQ ID NO: 29). Computational analysis recovered a complementary transcript belonging to Mink, a new member of the germinal center kinase (GCK) family (SEQ ID NO: 30) [Dan I. FEBS Lett. (2000)

469, 19-23] showing an overlap of at least 280 nucleotides to the AchRε mRNA, as schematically illustrated in Figure 16.

To validate the overlap of the two genes and to learn about their tissue distribution, northern analysis of a variety of human tissues was performed. Poly(A)-RNA containing membrane was hybridized with a 280 nucleotides RNA probes, corresponding to the overlap region in either antisense or sense orientation (SEQ ID NOs: 31 and 32, respectively).

As is evident from Figure 17a an AChRε transcript was expressed as a predominant 4 Kb band and had the highest expression in the heart, kidney and brain while surprisingly only a limited expression was observed in the skeletal muscle.

Hybridization with a MINK specific RNA probe revealed a major transcript of about 5 Kb, in accordance with previous results [Dan I. FEBS Lett. (2000) 469, 19-23] (Figure 17b). The mRNA transcript was ubiquitously expressed with strongest expression found in brain, liver, thymus, spleen and pancreas, again in agreement with Dan I. et al.

Conclusion:

The finding that AChRε and Mink genes are antisense each to one another with a significant overlap, and the fact that the two genes are co-expressed in some tissues (eg., brain) suggest the possibility that one of them may regulate the other under certain conditions.

EXAMPLE 5

Expression of Cyclin E2 mRNA and complementary transcripts in a variety of human cell-lines

Background:

The human cyclin E2 gene encodes a 404-amino-acid protein that is most closely related to cyclin E. Cyclin E2 associates with Cdk2 in a functional kinase complex that is inhibited by both p27(Kip1) and p21(Cip1). The catalytic activity associated with cyclin E2 complexes is cell cycle regulated and peaks at the G1/S transition. Overexpression of cyclin E2 in mammalian cells accelerates cell-cycle progression. Unlike cyclin E1, cyclin E2 levels are low to undetectable in

71

nontransformed cells and increase significantly in tumor-derived cells suggesting specific mechanism of regulation.

Results:

One RNA transcript was found for cyclin E2 (SEQ ID NO: 33. Computational search for natural antisense transcripts has revealed one complementary transcript for cyclin E2 messenger RNA (SEQ ID NO: 34). The overlap between the cyclin E2 sense RNA and the antisense transcript is of at least 72 nucleotides.

To confirm expression of the computationally retrieved antisense transcript for cyclin E2 as well as of cyclin E2 mRNA in human cell lines, two RNA-probes of 800 ribonucleotides were generated. Schematic location of the probes used for sense and antisense validation (SEQ ID NO: 44, Riboprobe#1 is illustrated in Figure 18).

As shown in Figure 19a, Riboprobe#1 detected two transcripts of approximately 3 Kb and 4.3 Kb. The absolute levels of the transcripts were quite heterogenic in all cell-lines examined. Both transcripts were completely absent from the Ln Cap cell line, while significantly high expression was observed in MCF-7 and DLD-1 lines, especially of the short transcript.

The same membrane was used to perform northern analysis with Riboprobe#2 in order to validate expression of antisense transcripts of cyclin E2. As is evident from Figure 19b, an antisense transcript 3.8 Kb long was observed in most cells assayed. Significantly high pattern of expression was observed in K562, MCF-7 and DLD-1 cell lines, while only a very moderate level of expression was detected in Ln Cap and HepG2 cell lines.

72

EXAMPLE 6

Co-regulated expression of CIDE-B and its complementary transcript upon induction of apoptosis

The discovery of a novel naturally occurring antisense transcript to the
apoptosis inducing factor, CIDE-B (see Example 2 hereinabove), suggested that the
latter may be regulated by its complementary transcript, thereby establishing a novel
mechanism of regulation. To address this, differential expression analysis of CIDE-B
expression and its endogenous antisense transcript expression was performed
following induction of apoptosis.

Materials and methods

Induction of apoptosis and reverse transcription analysis -
Monolayers of 293 cells were either left untreated (UT) or incubated with
increasing concentrations of etoposide or staurosporine (Sigma IL). Twenty-four
hours following addition of the drug, total RNA was extracted as decribed
hereinabove. Purified RNA was further treated with DNaseI. A reverse transcription
reaction were carried out with equivalent amounts of RNA in a final volume of 20 μl
containing 100 pmol of the oligo(dT) primer, 250 ng of total RNA, 0.5 mM each of
four deoxynucleoside triphosphates and 5 units of reverse transcriptase. The reaction
mixture was incubated at 65 °C for 5 min, 42 °C for 50 min and 70 °C for 15 min.
PCR was carried out in a final volume of 25 μl containing 12.5 pmol each of the
oligonucleotide primers derived of exons 3 and 7 of CIDE-B (SEQ ID NOs: 39 and
40), 1 μl of RT solution and 1.75 units of *Taq* polymerase. Amplification was carried
out by an initial denaturation step at 94 °C for 5 min followed by 35 cycles of [94 °C
for 30 s, 68 °C for 30 s, and 68 °C for 130 min]. At the end of the PCR amplification,
products were analyzed on agarose gels stained with ethidium bromide and visualized
with UV light.

Results

Amplification reaction yielded two major PCR products of 740 bp and 2285
bp (Figure 20). The small (740 bp) PCR product derived from the sense (CIDE-B)
strand, whereas the larger (2285 bp) product represented an intronless antisense
transcript. Evidently, an increase of sense transcript, concomitant with a decrease of
antisense transcript, was observed following treatment with etoposide (lanes 1-4) as

73

compared to untreated cells (lane 9), while no change was detected following staurosporine treatment (lanes 5-8).

These results suggest that following induction of apoptosis, antisense regulation of CIDE-B is abolished thereby allowing CIDE-B mediated apoptosis to proceed.

EXAMPLE 7

Reciprocal variation in sense and antisense expression of mouse nicotinic acetylcholine receptor, epsilon subunit during differentiation

The mouse nicotinic acetylcholine receptor, epsilon (mAchRε) subunit (SEQ ID NO: 35) has a critical function in a variety of differentiation processes. To address a novel concept of antisense regulation of AchRε-mediated differentiation, expression patterns of AchRε and its naturally occurring antisense transcript (SEQ ID NO: 36) were examined following induction of differentiation.

Materials and methods

Induction of apoptosis and reverse transcription analysis - C2 mouse myoblast cells were incubated with a differentiation medium (Dulbecco's modified Eagle's medium (DMEM) including 10 μg/ml insulin and 10 μ g/ml transferrin) or control medium (untreated) for 48 and 72 hours. Total RNA was extracted from treated and control cells and reverse-transcribed. PCR was done using F4 and R3 primers, derived from exon numbers 10 and 12 (last exon, SEQ ID NOs: 41 and 42, respectively) of the mouse nicotinic acetylcholine receptor, epsilon subunit (mAChRε) and directed at detecting sense and antisense transcripts (see Figure 21a).

Results

Amplification reaction showed a gradual increase in AchRε transcript expression, concomitant with the differentiation state of the cells. A second amplification product, which corresponded to an unspliced transcript was seen in untreated cells and disappeared following induction of differentiation. This fragment corresponds to a putative antisense transcript of the AchRε, and may represent an alternative 3' UTR of the Mink gene , of which the known transcript terminates 400 bp downstream to AchRε (see Example 4). To overcome possible competition between the two transcripts, another PCR reaction was carried out using antisense

74

specific riboprobes F4 and R4 (SEQ ID NO: 43). Reverse transcription products of this amlification reaction showed a single band which corresponded to a naturally occurring antisense transcript of the AchRε. As expected this transcript disappeared following induction of differentiation.

5   These results imply inverse regulation of the AchRε and its naturally occurring antisense transcript, during muscle cells differentiation from myoblasts to myotubes. Regulation may proceed, possibly through complementation of the sense and antisense transcripts to form dsRNA which can serve as a substrate for double strand RNA processing enzymes such as RNase H.

10

EXAMPLE 8

A polynucleotide database of sequences corresponding to the naturally occurring antisense transcripts identified by the present invention and their complementary sense sequences

15   Naturally occurring antisense sequences identified according to the teachings of the present invention and their corresponding sense sequences are provided in CD-ROM1-3 enclosed herewith (CD content is further described hereinbelow). File content is described in file AS_patent_data_description.doc on CD-ROM1. Generally a "seq" text file contains the actual polynucleotide sequences; a "table" file contains

20   summarized data pertaining to each sense-antisense sequence pair; an "alignments" file contains sequence alignments of sense and antisense overlapping regions.

The respective polypeptides, encoded by the naturally occurring antisense polynucleotides of the present invention, are described in the file "pep_seqs_133" in the attached CD-ROM-3. It should be noted that not all nucleotide sequences have a

25   corresponding polypeptide sequence.

The correlation between the nucleic acid sequence and the respective polypeptide is described in the file "Trans2Pep" in the attached CD-ROM-3. Each line in the "Trans2Pep" file represents a transcript name followed by a respective polypeptide name (eg: T92147_5 T92147_P6)

30   Annotation of the polypeptides encoded by the naturally occurring antisense sequences of the present invention appear in the "annotations_133" file in the attached CD-ROM-3, and are further described in Example 10 below.

Version 133

6 files : *table_133, seqs_133, alignments_133, annotations_133, pep_seqs_133,*
Trans2Pep.

table is a list of 175,644 pairs of transcripts representing 6230 pairs of contigs.

Numbering :   m_n

m – contigs' pair number.

n – number of transcripts' pair that belongs to a pair of contigs.

(each pair of contigs is represented by one or more pairs of transcripts)

seqs contains 99,414 sequences of all the transcripts, ordered by pairs and numbered
according to the list in *table.*

alignments contains the alignment of each pair of overlapping transcripts – 175,644
alignments.

Trans2Pep – table of 28,775 transcripts from seqs_133 that correspond to 14,658
peptides.

pep_seqs_133 - 14,658 peptide sequences in Fasta format.

annotations_133 – 14,658 annotations of 14,658 peptides.

Version 125

3 files : *table_125, seqs_125, alignments_125.*

table is a list of 223,181 pairs of transcripts representing 4018 pairs of contigs.

Numbering :   m_n

m – contigs' pair number.

n – number of transcripts' pair that belongs to a pair of contigs.

(each pair of contigs is represented by one or more pairs of transcripts)

seqs contains 79,884 sequences of all the transcripts, ordered by pairs and numbered
according to the list in *table.*

alignments contains the alignment of each pair of overlapping transcripts – 223,181
alignments.

"Table S1" and "Table S2" are further described in Example 9.

Table 3 below exemplifies the format of the Tables provided in CD-ROMs 2
and 3. Each row represents a pair of transcripts. The columns of Table 3 represent

76

(from the left): the serial number of the pair, the name of the first transcript, its length in nucleotides, the name of the second transcript, its length in nucleotides, the number of base pairs that overlap between the two transcripts, offsets of overlap beginning at the first transcript, offsets of overlap beginning at the second transcript.

Table 3

Serial No.	First transcript	First transcript length (nt)	Second transcript	Second transcript length (nt)	Overlap length (nt)	Start of overlap in first / in second transcript
570_0	AV705532_0 (SEQ ID NO: 1)	190	Z44352_15 (SEQ ID NO: 2)	783	OL: 52	OF1: 1 OF2: 1
570_1	AV705532_0	190	Z44352_14 (SEQ ID NO: 3)	1649	OL: 52	OF1: 1 OF2: 1
570_2	AV705532_0	190	Z44352_13 (SEQ ID NO: 4)	1861	OL: 52	OF1: 1 OF2: 1
571_0	AW070860_0 (SEQ ID NO: 5)	214	T81142_7 (SEQ ID NO: 6)	1934	OL: 54	OF1: 1 OF2: 1162
571_1	AW070860_0	214	T81142_6 (SEQ ID NO: 7)	2353	OL: 54	OF1: 1 OF2: 1162
571_2	AW070860_0	214	T81142_4 (SEQ ID NO: 8)	2500	OL: 54	OF1: 1 OF2: 1264
571_3	AW070860_0	214	T81142_3 (SEQ ID NO: 9)	947	OL: 54	OF1: 1 OF2: 171
571_4	AW070860_0	214	T81142_2 (SEQ ID NO: 10)	1366	OL: 54	OF1: 1 OF2: 171
572_0	BE046369_0 (SEQ ID NO: 11)	422	W26553_3 (SEQ ID NO: 12)	1532	OL: 52	OF1: 1 OF2: 1532
572_1	BE046369_0	422	W26553_2 (SEQ ID NO: 13)	1753	OL: 52	OF1: 1 OF2: 1753
572_2	BE046369_0	422	W26553_1 (SEQ ID NO: 14)	1832	OL: 52	OF1: 1 OF2: 1832

Pairs of transcripts are numbered, (within a contig pair, right to the underscore) that belong to a pair of contigs (numbered left to the underscore). Transcript names are arbitrary designataions.

Sequence alignment of the overlapping region in each sense and antisense pair of Table 1 is demonstrated in Figure 4a-k. Alignments were performed using the BLAST sequence alignment algorithm (Basic Local Alignment Search Tool, available through www.ncbi.nlm.nih.gov/BLAST). Interestingly, alignment profile shows high

level of variability with regard to overlap lengths. It is conceivable that short overlaps are due to technical reasons associated with insufficient sequence data.

The putative naturally occurring antisense transcripts identified by the present invention and disclosed in the enclosed CD-ROMs can be used to detect and/or treat a

5   variety of diseases, disorders or conditions, examples of which are listed hereinunder. For example, antisense transcripts or sequence information derived therefrom can be used to construct microarray kits (described in details in the preferred embodiments section) dedicated to diagnosing specific diseases, disorders or conditions.

The following sections list examples of proteins (subsection i), based on their

10   molecular function, which participate in variety of diseases (listed in subsection ii), which diseases can be diagnosed/treated using information derived from naturally occurring antisense transcripts such as those uncovered by the present invention.

The present invention is of biomolecular sequences, which can be classified to functional groups based on known activity of homologous sequences. This functional

15   group classification, allows the identification of diseases and conditions, which may be diagnosed and treated based on the novel sequence information and annotations of the present invention.

This functional group classification includes the following groups:

Proteins involved in Drug-Drug interactions:

20   The phrase "proteins involved in drug-drug interactions" refers to proteins involved in a biological process which mediates the interaction between at least two consumed drugs.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

25   altering expression of such proteins, may be used to modulate drug-drug interactions. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such drug-drug interactions.

Examples of these conditions include, but are not limited to the cytochrom P450

30   protein family, which is involved in the metabolism of many drugs. Examples of proteins, which are involved in drug-drug interactions are presented in Table 9.

Proteins involved in the metabolism of a pro-drug to a drug:

78

The phrase "proteins involved in the metabolism of a pro-drug to a drug" refers to proteins that activate an inactive pro-drug by chemically chaining it into a biologically active compound. Preferably, the metabolizing enzyme is expressed in the target tissue thus reducing systemic side effects.

5   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to modulate the metabolism of a pro-drug into drug. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used
10   for diagnosis of such conditions.

Examples of these proteins include, but are not limited to esterases hydrolyzing the cholesterol lowering drug simvastatin into its hydroxy acid active form.

MDR proteins:

15   The phrase "MDR proteins" refers to Multi Drug Resistance proteins that are responsible for the resistance of a cell to a range of drugs, usually by exporting these drugs outside the cell. Preferably, the MDR proteins are ABC binding cassette proteins. Preferably, drug resistance is associated with resistance to chemotherapy.

Pharmaceutical compositions including such proteins or protein encoding
20   sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transport of molecules and macromolecules such as neurotransmitters, hormones, sugar etc. is abnormal leading to various pathologies. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or
25   protein encoding sequences may be used for diagnosis of such diseases.

Examples of these proteins include, but are not limited to the multi-drug resistant transporter MDR1/P-glycoprotein, the gene product of MDR1, which belongs to the ATP-binding cassette (ABC) superfamily of membrane transporters and increases the resistance of malignant cells to therapy by exporting the therapeutic
30   agent out of the cell.

Hydrolases acting on amino acids:

79

The phrase "hydrolases acting on amino acids" refers to hydrolases acting on a pair of amino acids.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transfer of a glycosyl chemical group from one molecule to another is abnormal thus, a beneficial effect may be achieved by modulation of such reaction. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to reperfusion of clotted blood vessels by TPA (Tissue Plasminogen Activator) which converts the abundant, but inactive, zymogen plasminogen to plasmin by hydrolyzing a single ARG-VAL bond in plasminogen.

Transaminases:

The term "transaminases" refers to enzymes transferring an amine group from one compound to another.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transfer of an amine group from one molecule to another is abnormal thus, a beneficial effect may be achieved by modulation of such reaction. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such transaminases include, but are not limited to two liver enzymes, frequently used as markers for liver function - SGOT (Serum Glutamic-Oxalocetic Transaminase - AST) and SGPT (Serum Glutamic-Pyruvic Transaminase - ALT).

Immunoglobulins:

The term "immunoglobulins" refers to proteins that are involved in the immune and complement systems such as antigens and autoantigens, immunoglobulins, MHC and HLA proteins and their associated proteins.

80

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving the immune system such as inflammation, autoimmune diseases, infectious diseases, and

5   cancerous processes. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases and molecules that may be target for diagnostics include, but are not limited to members of the complement family such as C3 and C4

10   that their blood level is used for evaluation of autoimmune diseases and allergy state and C1 inhibitor that its absence is associated with angioedema. Thus, new variants of these genes are expected to be markers for similar events. Mutation in variants of the complement family may be associated with other immunological syndromes, such as increased bacterial infection that is associated with mutation in C3. C1 inhibitor was

15   shown to provide safe and effective inhibition of complement activation after reperfused acute myocardial infarction and may reduce myocardial injury [Eur. Heart J. 2002, 23(21):1670-7], thus, its variant may have the same or improved effect.

Transcription factor binding:

The phrase "transcription factor binding" refers to proteins involved in

20   transcription process by binding to nucleic acids, such as transcription factors, RNA and DNA binding proteins, zinc fingers, helicase, isomerase, histones, and nucleases.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins may be used to treat diseases involving

25   transcription factors binding proteins. Such treatment may be based on transcription factor that can be used to for modulation of gene expression associated with the disease. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

30   Examples of such diseases include, but are not limited to breast cancer associated with ErbB-2 expression that was shown to be successfully modulated by a transcription factor [Proc. Natl. Acad. Sci. U S A. 2000, 97(4):1495-500]. Examples of

novel transcription factors used for therapeutic protein production include, but are not limited to those described for Erythropoietin production [J. Biol. Chem. 2000, 275(43):33850-60; J. Biol. Chem. 2000, 275(43):33850-60] and zinc fingers protein transcription factors (ZFP-TF) variants [J. Biol. Chem. 2000, 275(43):33850-60].

5   *Small GTPase regulatory/interacting proteins:*

The phrase "Small GTPase regulatory/interacting proteins" refers to proteins capable of regulating or interacting with GTPase such as RAB escort protein, guanyl-nucleotide exchange factor, guanyl-nucleotide exchange factor adaptor, GDP-dissociation inhibitor, GTPase inhibitor, GTPase activator, guanyl-nucleotide

10   releasing factor, GDP-dissociation stimulator, regulator of G-protein signaling, RAS interactor, RHO interactor, RAB interactor, and RAL interactor.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which G-

15   proteases mediated signal-transduction is abnormal, either as a cause, or as a result of the disease. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to diseases related to

20   prenylation. Modulation of prenylation was shown to affect therapy of diseases such as osteoporosis, ischemic heart disease, and inflammatory processes. Small GTPases regulatory/interacting proteins are major component in the prenylation post translation modification, and are required to the normal activity of prenylated proteins. Thus, their variants may be used for therapy of prenylation associated diseases.

25   *Calcium binding proteins:*

The phrase "calcium binding proteins" refers to proteins involve in calcium binding, preferably, calcium binding proteins, ligand binding or carriers, such as diacylglycerol kinase, Calpain, calcium-dependent protein serine/threonine phosphatase, calcium sensing proteins, calcium storage proteins.

30   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat calcium involved diseases.

82

Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to diseases related to hypercalcemia, hypertension, cardiovascular disease, muscle diseases, gastro-intestinal diseases, uterus relaxing, and uterus. An example for therapy use of calcium binding proteins variant may be treatment of emergency cases of hypercalcemia, with secreted variants of calcium storage proteins.

Oxidoreductase:

The term "oxidoreductase" refers to enzymes that catalyze the removal of hydrogen atoms and electrons from the compounds on which they act. Preferably, oxidoreductases acting on the following groups of donors: CH-OH, CH-CH, CH-NH2, CH-NH; oxidoreductases acting on NADH or NADPH, nitrogenous compounds, sulfur group of donors, heme group, hydrogen group, diphenols and related substances as donors; oxidoreductases acting on peroxide as acceptor, superoxide radicals as acceptor, oxidizing metal ions, CH2 groups; oxidoreductases acting on reduced ferredoxin as donor; oxidoreductases acting on reduced flavodoxin as donor; and oxidoreductases acting on the aldehyde or oxo group of donors.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases caused by abnormal activity of oxidoreductases. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to malignant and autoimmune diseases in which the enzyme DHFR (DiHydroFolateReductase) that participates in folate metabolism and essential for *de novo* glycine and purine synthesis is the target for the widely used drug Methotrexate (MTX).

Receptors:

The term "receptors" refers to protein-binding sites on a cell's surface or interior, that recognize and binds to specific messenger molecule leading to a biological response,

83

such as signal transducers, complement receptors, ligand-dependent nuclear receptors, transmembrane receptors, GPI-anchored membrane-bound receptors, various coreceptors, internalization receptors, receptors to neurotransmitters, hormones and various other effectors and ligands.

5   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases caused by abnormal activity of receptors, preferably, receptors to neurotransmitters, hormones and various other effectors and ligands. Antibodies and polynucleotides such as PCR primers and

10   molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

        Examples of such diseases include, but are not limited to, chronic myelomonocytic leukemia caused by growth factor β receptor deficiency [Rao D. S., et al., (2001) Mol. Cell Biol., 21(22):7796-806], thrombosis associated with protease-

15   activated receptor deficiency [Sambrano G. R., et al., (2001) Nature, 413(6851):26-7], hypercholesterolemia associated with low density lipoprotein receptor deficiency [Koivisto U. M., et al., (2001) Cell, 105(5):575-85], familial Hibernian fever associated with tumor necrosis factor receptor deficiency [Simon A., et al., (2001) Ned Tijdschr Geneeskd, 145(2):77-8], colitis associated with immunoglobulin E receptor expression

20   [Dombrowicz D., et al., (2001) J. Exp. Med., 193(1):25-34], and alagille syndrome associated with Jagged1 [Stankiewicz P. et al., (2001) Am. J. Med. Genet., 103(2):166-71], breast cancer associated with mutated BRCA2 and androgen. Therapeutic applications of nuclear receptors variants may be based on secreted version of receptors such as the thyroid nuclear receptor that by binding plasma free thyroid hormone to

25   reduce its levels may have a therapeutic effect in cases of thyrotoxicosis. A secreted version of glucocorticoid nuclear receptor, by binding plasma free cortisol, thus, reducing, may have a therapeutic effect in cases of Cushing's disease (a disease associated with high cortisole levels in the plasma).

        Another example of a secreted variant of a receptor is a secreted form of the TNF

30   receptor, which is used to treat conditions in which reduction of TNF levels is of benefit including Rheumatoid Arthritis, Juvenile Rheumatoid Arthritis, Psoriatic Arthritis and Ankylosing Spondylitis.

Protein serine/threonine kinases:

The phrase "protein serine/threonine kinases" refers to proteins which phosphorylate serine/threonine residues, mainly involved in signal transduction, such as transmembrane receptor protein serine/threonine kinase, 3-phosphoinositide-dependent protein kinase, DNA-dependent protein kinase, G-protein-coupled receptor phosphorylating protein kinase, SNF1A/AMP-activated protein kinase, casein kinase, calmodulin regulated protein kinase, cyclic-nucleotide dependent protein kinase, cyclin-dependent protein kinase, eukaryotic translation initiation factor 2α kinase, galactosyltransferase-associated kinase, glycogen synthase kinase 3, protein kinase C, receptor signaling protein serine/threonine kinase, ribosomal protein S6 kinase, and IkB kinase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases ameliorated by a modulating kinase activity. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to schizophrenia. 5-HT(2A) serotonin receptor is the principal molecular target for LSD-like hallucinogens and atypical antipsychotic drugs. It has been shown that a major mechanism for the attenuation of this receptor signaling following agonist activation typically involves the phosphorylation of serine and/or threonine residues by various kinases. Therefore, serine/threonine kinases specific for the 5-HT(2A) serotonin receptor may serve as drug targets for a disease such as schizophrenia. Other diseases that may be treated through serine/thereonine kinases modulation are Peutz-Jeghers syndrome (PJS, a rare autosomal-dominant disorder characterized by hamartomatous polyposis of the gastrointestinal tract and melanin pigmentation of the skin and mucous membranes [*Hum. Mutat.* 2000, 16(1):23-30], breast cancer [Oncogene. 1999, 18(35):4968-73], Type 2 diabetes insulin resistance [Am. J. Cardiol. 2002, 90(5A):11G-18G], and fanconi anemia [Blood. 2001, 98(13):3650-7].

Channel/pore class transporters:

The phrase "Channel/pore class transporters" refers to proteins that mediate the transport of molecules and macromolecules across membranes, such as α-type channels, porins, and pore-forming toxins.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transport of molecules and macromolecules are abnormal, therefore leading to various pathologies. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to diseases of the nerves system such as Parkinson, diseases of the hormonal system, diabetes and infectious diseases such as bacterial and fungal infections. For example, α-hemolysin, is a protein product of S. aureus which creates ion conductive pores in the cell membrane, thereby deminishing its integrity.

Hydrolases, acting on acid anhydrides:

The phrase "hydrolases, acting on acid anhydrides" refers to hydrolytic enzymes that are acting on acid anhydrides, such as hydrolases acting on acid anhydrides in phosphorus-containing anhydrides or in sulfonyl-containing anhydrides, hydrolases catalyzing transmembrane movement of substances, and involved in cellular and subcellular movement.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins may be used to treat diseases in which the hydrolase-related activities are abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to glaucoma treated with carbonic anhydrase inhibitors (e.g. Dorzolamide), peptic ulcer disease treated with $H^{(+)}K^{(+)}ATPase$ inhibitors that were shown to affect disease by blocking gastric carbonic anhydrase (e.g. Omeprazole).

86

Transferases, transferring phosphorus-containing groups:

The phrase "transferases, transferring phosphorus-containing groups " refers to enzymes that catalyze the transfer of phosphate from one molecule to another, such as phosphotransferases using the following groups as acceptors: alcohol group, carboxyl

5   group, nitrogenous group, phosphate; phosphotransferases with regeneration of donors catalyzing intramolecular transfers; diphosphotransferases; nucleotidyltransferase; and phosphotransferases for other substituted phosphate groups.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

10 altering expression of such proteins may be used to treat diseases in which the transfer of a phosphorous containing functional group to a modulated moiety is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

15 Examples of such diseases include, but are not limited to acute MI [Ann. Emerg. Med. 2003, 42(3):343-50], Cancer [Oral. Dis. 2003, 9(3):119-28; J. Surg. Res. 2003, 113(1):102-8] and Alzheimer's disease [Am. J. Pathol. 2003, 163(3):845-58]. Examples for possible utilities of such transferases for drug improvement include, but are not limited to aminoglycosides treatment (antibiotics) to which resistance is

20 mediated by aminoglycoside phosphotransferases [Front. Biosci. 1999, 1;4:D9-21]. Using aminoglycoside phosphotransferases variants or inhibiting these enzymes may reduce aminoglycosides resistance. Since aminoglycosides can be toxic to some patients, proving the expression of aminoglycoside phosphotransferases in a patient can deter from treating him with aminoglycosides and risking the patient in vain.

25 *Phosphoric monoester hydrolases:*

The phrase "phosphoric monoester hydrolases" refers to hydrolytic enzymes that are acting on ester bonds, such as nuclease, sulfuric ester hydrolase, carboxylic ester hydrolase, thiolester hydrolase, phosphoric monoester hydrolase, phosphoric diester hydrolase, triphosphoric monoester hydrolase, diphosphoric monoester

30 hydrolase, and phosphoric triester hydrolase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

altering expression of such proteins, may be used to treat diseases in which the hydrolytic cleavage of a covalent bond with accompanying addition of water (-H being added to one product of the cleavage and –OH to the other), is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify
5   such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to diabetes and CNS diseases such as Parkinson and cancer.

Enzyme inhibitors:

10   The term "enzyme inhibitors" refers to inhibitors and suppressors of other proteins and enzymes, such as inhibitors of: kinases, phosphatases, chaperones, guanylate cyclase, DNA gyrase, ribonuclease, proteasome inhibitors, diazepam-binding inhibitor, ornithine decarboxylase inhibitor, GTPase inhibitors, dUTP pyrophosphatase inhibitor, phospholipase inhibitor, proteinase inhibitor, protein
15   biosynthesis inhibitors, and α-amylase inhibitors.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which beneficial effect may be achieved by modulating the activity of inhibitors and suppressors of
20   proteins and enzymes. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to α-1 antitrypsin (a natural serine proteases, which protects the lung and liver from proteolysis) deficiency
25   associated with emphysema, COPD and liver chirosis. α-1 antitrypsin is also used for diagnostics in cases of unexplained liver and lung disease. A variant of this enzyme may act as protease inhibitor or a diagnostic target for related diseases.

Electron transporters:

The term "Electron transporters" refers to ligand binding or carrier proteins
30   involved in electron transport such as flavin-containing electron transporter, cytochromes, electron donors, electron acceptors, electron carriers, and cytochrome-c oxidases.

88

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which beneficial effect may be achieved by modulating the activity of electron transporters. Antibodies

5   and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to cyanide toxicity, resulting from cyanide binding to ubiquitous metalloenzymes rendering them inactive,

10   and interfering with the electron transport. Novel electron transporters to which cyanide can bind may serve as drug targets for new cyanide antidotes.

Transferases, transferring glycosyl groups:

The phrase "transferases, transferring glycosyl groups" refers to enzymes that catalyze the transfer of a glycosyl chemical group from one molecule to another such

15   as murein lytic endotransglycosylase E, and sialyltransferase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transfer of a glycosyl chemical group is abnormal. Antibodies and polynucleotides such as

20   PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Ligases, forming carbon-oxygen bonds:

The phrase "ligases, forming carbon-oxygen bonds" refers to enzymes that catalyze the linkage between carbon and oxygen such as ligase forming aminoacyl-

25   tRNA and related compounds.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the linkage between carbon and oxygen in an energy dependent process is abnormal. Antibodies

30   and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

89

Ligases:

The term "ligases" refers to enzymes that catalyze the linkage of two molecules, generally utilizing ATP as the energy donor, also called synthetase. Examples for ligases are enzymes such as β-alanyl-dopamine hydrolase, carbon-oxygen bonds forming ligase, carbon-sulfur bonds forming ligase, carbon-nitrogen bonds forming ligase, carbon-carbon bonds forming ligase, and phosphoric ester bonds forming ligase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the joining together of two molecules in an energy dependent process is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to neurological disorders such as Parkinson's disease [Science. 2003, 302(5646):819-22; J. Neurol. 2003, 250 Suppl. 3:III25-III29] or epilepsy [Nat. Genet. 2003, 35(2):125-7], cancerous diseases [Cancer Res. 2003, 63(17):5428-37; Lab. Invest. 2003, 83(9):1255-65], renal diseases [Am. J. Pathol. 2003, 163(4):1645-52], infectious diseases [Arch. Virol. 2003, 148(9):1851-62] and fanconi anemia [Nat. Genet. 2003, 35(2):165-70].

Hydrolases, acting on glycosyl bonds:

The phrase "hydrolases, acting on glycosyl bonds" refers to hydrolytic enzymes that are acting on glycosyl bonds such as hydrolases hydrolyzing N-glycosyl compounds, S-glycosyl compounds, and O-glycosyl compounds.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the hydrolase-related activities are abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include cancerous diseases [J. Natl. Cancer Inst. 2003, 95(17):1263-5; Carcinogenesis. 2003, 24(7):1281-2; author reply 1283] vascular diseases [J. Thorac. Cardiovasc. Surg. 2003, 126(2):344-57], gastrointestinal diseases

90

such as colitis [J. Immunol. 2003, 171(3):1556-63] or liver fibrosis [World J. Gastroenterol. 2002, 8(5):901-7].

Kinases:

The term "kinases" refers to enzymes which phosphorylate serine/threonine or
5   tyrosine residues, mainly involved in signal transduction. Examples for kinases include enzymes such as 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase, NAD($^+$) kinase, acetylglutamate kinase, adenosine kinase, adenylate kinase, adenylsulfate kinase, arginine kinase, aspartate kinase, choline kinase, creatine kinase, cytidylate kinase, deoxyadenosine kinase, deoxycytidine
10   kinase, deoxyguanosine kinase, dephospho-CoA kinase, diacylglycerol kinase, dolichol kinase, ethanolamine kinase, galactokinase, glucokinase, glutamate 5-kinase, glycerol kinase, glycerone kinase, guanylate kinase, hexokinase, homoserine kinase, hydroxyethylthiazole kinase, inositol/phosphatidylinositol kinase, ketohexokinase, mevalonate kinase, nucleoside-diphosphate kinase, pantothenate kinase,
15   phosphoenolpyruvate carboxykinase, phosphoglycerate kinase, phosphomevalonate kinase, protein kinase, pyruvate dehydrogenase (lipoamide) kinase, pyruvate kinase, ribokinase, ribose-phosphate pyrophosphokinase, selenide, water dikinase, shikimate kinase, thiamine pyrophosphokinase, thymidine kinase, thymidylate kinase, uridine kinase, xylulokinase, 1D-myo-inositol-trisphosphate 3-kinase, phosphofructokinase,
20   pyridoxal kinase, sphinganine kinase, riboflavin kinase, 2-dehydro-3-deoxygalactonokinase, 2-dehydro-3-deoxygluconokinase, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, GTP pyrophosphokinase, L-fuculokinase, L-ribulokinase, L-xylulokinase, isocitrate dehydrogenase (NADP$^+$) kinase, acetate kinase, allose kinase, carbamate kinase, cobinamide kinase, diphosphate-purine nucleoside kinase,
25   fructokinase, glycerate kinase, hydroxymethylpyrimidine kinase, hygromycin-B kinase, inosine kinase, kanamycin kinase, phosphomethylpyrimidine kinase, phosphoribulokinase, polyphosphate kinase, propionate kinase, pyruvate,water dikinase, rhamnulokinase, tagatose-6-phosphate kinase, tetraacyldisaccharide 4'-kinase, thiamine-phosphate kinase, undecaprenol kinase, uridylate kinase, N-
30   acylmannosamine kinase, D-erythro-sphingosine kinase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

altering expression of such proteins, may be used to treat diseases which may be ameliorated by a modulating kinase activity. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, acute lymphoblastic leukemia associated with spleen tyrosine kinase deficiency [Goodman P.A., et al., (2001) Oncogene, 20(30):3969-78], ataxia telangiectasia associated with ATM kinase deficiency [Boultwood J., (2001) J. Clin. Pathol., 54(7):512-6], congenital haemolytic anaemia associated with erythrocyte pyruvate kinase deficiency [Zanella A., et al., (2001) Br. J. Haematol., 113(1):43-8], mevalonic aciduria caused by mevalonate kinase deficiency [Houten S. M., et al., (2001) Eur. J. Hum. Genet., 9(4):253-9], and acute myelogenous leukemia associated with over-expressed death-associated protein kinase [Guzman M. L., et al., (2001) Blood, 97(7):2177-9].

Nucleotide binding:

The term "nucleotide binding" refers to ligand binding or carrier proteins, involved in physical interaction with a nucleotide, preferably, any compound consisting of a nucleoside that is esterified with [ortho]phosphate or an oligophosphate at any hydroxyl group on the glycose moiety, such as purine nucleotide binding proteins.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases that are associated with abnormal nucleotide binding. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to Gout (a syndrome characterized by high urate level in the blood). Since urate is a breakdown metabolite of purines, reducing purines serum levels could have a therapeutic effect in Gout disease.

Tubulin binding:

The term "tubulin binding" refers to binding proteins that bind tubulin such as microtubule binding proteins.

92

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are associated with abnormal tubulin activity or structure. Binding the products of the genes of this

5   family, or antibodies reactive therewith, can modulate a plurality of tubulin activities as well as change microtubulin structure. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, Alzheimer's disease

10   associated with t-complex polypeptide 1 deficiency [Schuller E., et al., (2001) Life Sci., 69(3):263-70], neurodegeneration associated with *apo*E deficiency [Masliah E., et al., (1995) Exp. Neurol., 136(2):107-22], progressive axonopathy associated with disfuctional neurofilaments [Griffiths I. R., et al., (1989) Neuropathol. Appl. Neurobiol., 15(1):63-74], familial frontotemporal dementia associated with tau deficiency [astor P., et al., (2001)

15   Ann. Neurol., 49(2):263-7], and colon cancer suppressed by APC [White R. L., (1997) Pathol. Biol. (Paris), 45(3):240-4]. En example for a drug whose target is tubulin is the anticancer drug - Taxol. Drugs having similar mechanism of action (interfering with tubulin polymerization) may be developed based on tubulin binding proteins.

Receptor signaling proteins:

20   The phrase "receptor signaling proteins" refers to receptor proteins involved in signal transduction such as receptor signaling protein serine/threonine kinase, receptor signaling protein tyrosine kinase, receptor signaling protein tyrosine phosphatase, aryl hydrocarbon receptor nuclear translocator, hematopoeitin/interferon-class (D200-domain) cytokine receptor signal transducer, transmembrane receptor protein tyrosine

25   kinase signaling protein, transmembrane receptor protein serine/threonine kinase signaling protein, receptor signaling protein serine/threonine kinase signaling protein, receptor signaling protein serine/threonine phosphatase signaling protein, small GTPase regulatory/interacting protein, receptor signaling protein tyrosine kinase signaling protein, and receptor signaling protein serine/threonine phosphatase.

30   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the signal-

93

transduction is abnormal, either as a cause, or as a result of the disease. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, complete hypogonadotropic hypogonadism associated with GnRH receptor deficiency [Kottler M. L., et a., (2000) J. Clin. Endocrinol. Metab., 85(9):3002-8], severe combined immunodeficiency disease associated with IL-7 receptor deficiency [Puel A. and Leonard W. J., (2000) Curr. Opin. Immunol., 12(4):468-7], schizophrenia associated N-methyl-D-aspartate receptor deficiency [Mohn A.R., et al., (1999) Cell, 98(4):427-36], Yesinia-associated arthritis associated with tumor necrosis factor receptor p55 deficiency [Zhao Y. X., et al., (1999) Arthritis Rheum., 42(8):1662-72], and Dwarfism of Sindh caused by growth hormone-releasing hormone receptor deficiency [aheshwari H. G., et al., (1998) J. Clin. Endocrinol. Metab., 83(11):4065-74].

Molecular function unknown:

The phrase "molecular function unknown" refers to various proteins with unknown molecular function, such as cell surface antigens.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which regulation of the recognition, or participation or bind of cell surface antigens to other moieties may have therapeutic effect. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, autoimmune diseases, various infectious diseases, cancer diseases which involve non cell surface antigens recognition and activity.

Enzyme activators:

The term "enzyme activators" refers to enzyme regulators such as activators of: kinases, phosphatases, sphingolipids, chaperones, guanylate cyclase, tryptophan hydroxylase, proteases, phospholipases, caspases, proprotein convertase 2 activator, cyclin-dependent protein kinase 5 activator, superoxide-generating NADPH oxidase

94

activator, sphingomyelin phosphodiesterase activator, monophenol monooxygenase activator, proteasome activator, and GTPase activator.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

5   altering expression of such proteins, may be used to treat diseases in which beneficial effect may be achieved by modulating the activity of activators of proteins and enzymes. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

10   Examples of such diseases include, but are not limited to all complement related diseases, as most complement proteins activate by cleavage other complement proteins.

Transferases, transferring one-carbon groups:

The phrase "transferases, transferring one-carbon groups" refers enzymes that catalyze the transfer of a one-carbon chemical group from one molecule to another

15   such as methyltransferase, amidinotransferase, hydroxymethyl-, formyl- and related transferase, carboxyl- and carbamoyltransferase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transfer

20   of a one-carbon chemical group from one molecule to another is abnormal so that a beneficial effect may be achieved by modulation of such reaction. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Transferases:

25   The term "transferases" refers to enzymes that catalyze the transfer of a chemical group, preferably, a phosphate or amine from one molecule to another. It includes enzymes such as transferases, transferring one-carbon groups, aldehyde or ketonic groups, acyl groups, glycosyl groups, alkyl or aryl (other than methyl) groups, nitrogenous,   phosphorus-containing   groups,   sulfur-containing   groups,

30   lipoyltransferase, deoxycytidyl transferases.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of

altering expression of such proteins, may be used to treat diseases in which the transfer of a chemical group from one molecule to another is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

5   Examples of such diseases include, but are not limited to cancerous diseases such as prostate cancer [Urology. 2003, 62(5 Suppl 1):55-62] or lung cancer [Invest. New Drugs. 2003, 21(4):435-43; JAMA. 2003, 22;290(16):2149-58], psychiatric disorders [Am. J. Med. Genet. 2003, 15;123B(1):64-9], colorectal disease such as Crohn's disease [Dis. Colon Rectum. 2003, 46(11):1498-507] or celiac diseases [N

10   Engl. J. Med. 2003, 349(17):1673-4; author reply 1673-4], neurological diseases such as Prkinson's disease [J. Chem Neuroanat. 2003, 26(2):143-51], Alzheimer disease [Hum. Mol. Genet. 2003 21] or Charcot-Marie-Tooth Disease [Mol. Biol. Evol. 2003 31].

Chaperones:

15   The term "chaperones" refers to functional classes of unrelated families of proteins that assist the correct non-covalent assembly of other polypeptide-containing structures *in vivo*, but are not components of these assembled structures when they a performing their normal biological function. The group of chaperones include proteins such as ribosomal chaperone, peptidylprolyl isomerase, lectin-binding chaperone,

20   nucleosome assembly chaperone, chaperonin ATPase, cochaperone, heat shock protein, HSP70/HSP90 organizing protein, fimbrial chaperone, metallochaperone, tubulin folding, and HSC70-interacting protein.

        Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering

25   expression of such proteins, may be used to treat diseases which are associated with abnormal protein activity, structure, degradation or accumulation of proteins. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

        Examples of such diseases include, but are not limited to neurological syndromes

30   [J. Neuropathol. Exp. Neurol. 2003, 62(7):751-64; Antioxid Redox Signal. 2003, 5(3):337-48; J. Neurochem. 2003, 86(2):394-404], neurological diseases such as Parkinson's disease [Hum. Genet. 2003, 6; Neurol Sci. 2003, 24(3):159-60; J. Neurol.

2003, 250 Suppl. 3:III25-III29] ataxia [J. Hum. Genet. 2003;48(8):415-9] or Alzheimer diseases [J. Mol. Neurosci. 2003, 20(3):283-6; J. Alzheimers Dis. 2003, 5(3):171-7], cancerous diseases [Semin. Oncol. 2003, 30(5):709-16], prostate cancer [Semin. Oncol. 2003, 30(5):709-16] metabolic diseases [J Neurochem. 2003, 87(1):248-56], infectious diseases, such as prion infection [EMBO J. 2003, 22(20):5435-5445]. Chaperones may be also used for manipulating therapeutic proteins binding to their receptors therefore, improving their therapeutic effect.

Cell adhesion molecule:

The phrase "cell adhesion molecule" refers to proteins that serve as adhesion molecules between adjoining cells such as membrane-associated protein with guanylate kinase activity, cell adhesion receptor, neuroligin, calcium-dependent cell adhesion molecule, selectin, calcium-independent cell adhesion molecule, and extracellular matrix protein.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which adhesion between adjoining cells is involved, typically conditions in which the adhesion is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to cancer in which abnormal adhesion may cause and enhance the process of metastasis and abnormal growth and development of various tissues in which modulation adhesion among adjoining cells can improve the condition. Leucocyte-endothlial interactions characterized by adhesion molecules involved in interactions between cells lead to a tissue injury and ischemia reperfusion disorders in which activated signals generated during ischemia may trigger an exuberant inflammatory response during reperfusion, provoking greater tissue damage than initial ischemic insult [Crit. Care Med. 2002, 30(5 Suppl):S214-9]. The blockade of leucocyte-endothelial adhesive interactions has the potential to reduce vascular and tissue injury. This blockade may be achieved using a soluble variant of the adhesion molecule.

97

States of septic shock and ARDS involve large recruitment of neutrophil cells to the damaged tissues. Neutrophil cells bind to the endothelial cells in the target tissues through adhesion molecules. Neutrophils possess multiple effector mechanisms that can produce endothelial and lung tissue injury, and interfere with pulmonary gas transfer by disruption of surfactant activity [Eur. J. Surg. 2002, 168(4):204-14]. In such cases, the use of soluble variant of the adhesion molecule may decrease the adhesion of neutrophils to the damaged tissues.

Examples of such diseases include, but are not limited to, Wiskott-Aldrich syndrome associated with WAS deficiency [Westerberg L., et al., (2001) Blood, 98(4):1086-94], asthma associated with intercellular adhesion molecule-1 deficiency [Tang M. L. and Fiscus L. C., (2001) Pulm. Pharmacol. Ther., 14(3):203-10], intra-atrial thrombogenesis associated with increased von Willebrand factor activity [Fukuchi M., et al., (2001) J. Am. Coll. Cardiol., 37(5):1436-42], junctional epidermolysis bullosa associated with laminin 5-β-3 deficiency [Robbins P. B., et al., (2001) Proc. Natl. Acad. Sci., 98(9):5193-8], and hydrocephalus caused by neural adhesion molecule L1 deficiency [Rolf B., et al., (2001) Brain Res., 891(1-2):247-52].

Motor proteins:

The term "motor proteins" refers to proteins that generate force or energy by the hydrolysis of ATP and that function in the production of intracellular movement or transportation. Examples of such proteins include microfilament motor, axonemal motor, microtubule motor, and kinetochore motor (dynein, kinesin, or myosin).

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which force or energy generation is impaired. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, malignant diseases where microtubules are drug targets for a family of anticancer drugs such as myodystrophies and myopathies [Trends Cell Biol. 2002, 12(12):585-91], neurological disorders [Neuron. 2003, 25;40(1):25-40; Trends Biochem. Sci. 2003, 28(10):558-65;

98

Med. Genet. 2003, 40(9):671-5], and hearing impairment [Trends Biochem. Sci. 2003, 28(10):558-65].

Defense/immunity proteins:

The term "defense/immunity proteins" refers to proteins that are involved in 5 the immune and complement systems such as acute-phase response proteins, antimicrobial peptides, antiviral response proteins, blood coagulation factors, complement components, immunoglobulins, major histocompatibility complex antigens and opsonins.

Pharmaceutical compositions including such proteins or protein encoding 10 sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving the immunological system including inflammation, autoimmune diseases, infectious diseases, as well as cancerous processes or diseases which are manifested by abnormal coagulation processes, which may include abnormal bleeding or excessive coagulation. Antibodies and 15 polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, late (C5-9) complement component deficiency associated with opsonin receptor allotypes [Fijen C. A., et al., (2000) Clin. Exp. Immunol., 120(2):338-45], combined 20 immunodeficiency associated with defective expression of MHC class II genes [Griscelli C., et al., (1989) Immunodefic. Rev. 1(2):135-53], loss of antiviral activity of CD4 T cells caused by neutralization of endogenous TNFα [Pavic I., et al., (1993) J. Gen. Virol., 74 (Pt 10):2215-23], autoimmune diseases associated with natural resistance-associated macrophage protein deficiency [Evans C. A., et al., (2001) 25 Neurogenetics, 3(2):69-78], Epstein-Barr virus-associated lymphoproliferative disease inhibited by combined GM-CSF and IL-2 therapy [Baiocchi R. A., et al., (2001) J. Clin. Invest., 108(6):887-94], and sepsis in which activated protein C is a therapeutic protein itself.

Intracellular transporters:

30   The term "intracellular transporters" refers to proteins that mediate the transport of molecules and macromolecules inside the cell, such as intracellular

99

nucleoside transporter, vacuolar assembly proteins, vesicle transporters, vesicle fusion proteins, type II protein secretors.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering 5   expression of such proteins, may be used to treat diseases in which the transport of molecules and macromolecules is abnormal leading to various pathologies. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Transporters:

10   The term "transporters" refers to proteins that mediate the transport of molecules and macromolecules, such as channels, exchangers, and pumps. Transporters include proteins such as: amine/polyamine transporter, lipid transporter, neurotransmitter transporter, organic acid transporter, oxygen transporter, water transporter, carriers, intracellular transports, protein transporters, ion transporters, 15   carbohydrate transporter, polyol transporter, amino acid transporters, vitamin/cofactor transporters, siderophore transporter, drug transporter, channel/pore class transporter, group translocator, auxiliary transport proteins, permeases, murein transporter, organic alcohol transporter, nucleobase, nucleoside, and nucleotide and nucleic acid transporters.

20   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the transport of molecules and macromolecules such as neurotransmitters, hormones, sugar etc. is impaired leading to various pathologies. Antibodies and polynucleotides such as PCR 25   primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, glycogen storage disease caused by glucose-6-phosphate transporter deficiency [Hiraiwa H., and Chou J. Y. (2001) DNA Cell Biol., 20(8):447-53], tangier disease associated with ATP-binding 30   cassette transporter-1 deficiency [McNeish J., et al., (2000) Proc. Natl. Acad. Sci., 97(8):4245-50], systemic primary carnitine deficiency associated with organic cation transporter deficiency [Tang N. L., et al., (1999) Hum. Mol. Genet., 8(4):655-60], Wilson

100

disease associated with copper-transporting ATPases deficiency [Payne A. S., et al., (1998) Proc. Natl. Acad. Sci. 95(18):10854-9], and atelosteogenesis associated with diastrophic dysplasia sulphate transporter deficiency [Newbury-Ecob R., (1998) J. Med. Genet., 35(1):49-53], Central Nervous system diseases treated by inhibiting

5   neurotransmitter transporter (e.g. Depression, treated with serotonin transporters inhibitors – Prozac), and Cystic fibrosis mediated by the chloride channel CFTR. Other transporter related diseases are cancer [Oncogene. 2003, 22(38):6005-12] and especially cancer resistant to treatment [Oncologist. 2003, 8(5):411-24; J. Med. Invest. 2003, 50(3-4):126-35], infectious diseases, especially fungal infections [Annu. Rev. Phytopathol.

10   2003, 41:641-67], neurological diseases, such as Parkinson [FASEB J. 2003, Sep 4 [Epub ahead of print]], and cardiovascular diseases, including hypercholesterolemia [Am. J. Cardiol. 2003, 92(4B):10K-16K].

There are about 30 membrane transporter genes linked to a known genetic clinical syndrome. Secreted versions of splice variants of transporters may be therapeutic as the

15   case with soluble receptors. These transporters may have the capability to bind the compound in the serum they would normally bind on the membrane. For example, a secreted form ATP7B, a transporter involved in Wilson's disease, is expected to bind plasma Copper, therefore have a desired therapeutic effect in Wilson's disease.

Lyases:

20   The term "lyases" refers to enzymes that catalyze the formation of double bonds by removing chemical groups from a substrate without hydrolysis or catalyze the addition of chemical groups to double bonds. It includes enzymes such as carbon-carbon lyase, carbon-oxygen lyase, carbon-nitrogen lyase, carbon-sulfur lyase, carbon-halide lyase, and phosphorus-oxygen lyase.

25   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the double bonds formation catalyzed by these enzymes is impaired. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein

30   encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, autoimmune diseases [JAMA. 2003, 290(13):1721-8; JAMA. 2003, 290(13):1713-20], diabetes [Diabetes.

101

2003, 52(9):2274-8], neurological disorders such as epilepsy [J. Neurosci. 2003, 23(24):8471-9], Parkinson [J. Neurosci. 2003, 23(23):8302-9; Lancet. 2003, 362(9385):712] or Creutzfeldt-Jakob disease [Clin. Neurophysiol. 2003, 114(9):1724-8], and cancerous diseases [J. Pathol. 2003, 201(1):37-45; J. Pathol. 2003, 201(1):37-45; Cancer Res. 2003, 63(16):4952-9; Eur. J. Cancer. 2003, 39(13):1899-903].

Actin binding proteins:

The phrase "actin binding proteins" refers to proteins binding actin as actin cross-linking, actin bundling, F-actin capping, actin monomer binding, actin lateral binding, actin depolymerizing, actin monomer sequestering, actin filament severing, actin modulating, membrane associated actin binding, actin thin filament length regulation, and actin polymerizing proteins.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which actin binding is impaired. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, neuromuscular diseases such as muscular dystrophy [Neurology. 2003, 61(3):404-6], Cancerous diseases [Urology. 2003, 61(4):845-50; J. Cutan. Pathol. 2002, 29(7):430; Cancer. 2002, 94(6):1777-86; Clin. Cancer Res. 2001, 7(8):2415-24; Breast Cancer Res. Treat. 2001, 65(1):11-21], renal diseases such as glomerulonephritis [J. Am. Soc. Nephrol. 2002, 13(2):322-31; Eur. J. Immunol. 2001, 31(4):1221-7], and gastrointestinal diseases such as Crohn's disease [J. Cell Physiol. 2000, 182(2):303-9].

Protein binding proteins:

The phrase "protein binding proteins" refers to proteins involved in diverse biological functions through binding other proteins. Examples of such biological function include intermediate filament binding, LIM-domain binding, LLR-domain binding, clathrin binding, ARF binding, vinculin binding, KU70 binding, troponin C binding PDZ-domain binding, SH3-domain binding, fibroblast growth factor binding, membrane-associated protein with guanylate kinase activity interacting, Wnt-protein binding , DEAD/H-box RNA helicase binding, β-amyloid binding, myosin binding,

102

TATA-binding protein binding DNA topoisomerase I binding, polypeptide hormone binding, RHO binding, FH1-domain binding, syntaxin-1 binding, HSC70-interacting, transcription factor binding, metarhodopsin binding, tubulin binding, JUN kinase binding, RAN protein binding, protein signal sequence binding, importin α export

5   receptor, poly-glutamine tract binding, protein carrier, β-catenin binding, protein C-terminus binding, lipoprotein binding, cytoskeletal protein binding protein, nuclear localization sequence binding, protein phosphatase 1 binding, adenylate cyclase binding, eukaryotic initiation factor 4E binding, calmodulin binding, collagen binding, insulin-like growth factor binding, lamin binding, profilin binding, tropomyosin

10   binding, actin binding, peroxisome targeting sequence binding, SNARE binding, and cyclin binding.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are associated with

15   impaired protein binding. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, neurological and psychiatric diseases [J. Neurosci. 2003, 23(25):8788-99; Neurobiol. Dis. 2003, 14(1):146-

20   56; J. Neurosci. 2003, 23(17):6956-64; Am. J. Pathol. 2003, 163(2):609-19], and cancerous diseases [Cancer Res. 2003, 63(15):4299-304; Semin. Thromb. Hemost. 2003, 29(3):247-58; Proc. Natl. Acad. Sci. U S A. 2003, 100(16):9506-11].

Ligand binding or carrier proteins:

The phrase "ligand binding or carrier proteins" refers to proteins involved in

25   diverse biological functions such as: pyridoxal phosphate binding, carbohydrate binding, magnesium binding, amino acid binding, cyclosporin A binding, nickel binding, chlorophyll binding, biotin binding, penicillin binding, selenium binding, tocopherol binding, lipid binding, drug binding, oxygen transporter, electron transporter, steroid binding, juvenile hormone binding, retinoid binding, heavy metal

30   binding, calcium binding, protein binding, glycosaminoglycan binding, folate binding, odorant binding, lipopolysaccharide binding and nucleotide binding.

103

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are associated with impaired function of these proteins. Antibodies and polynucleotides such as PCR primers

5   and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, neurological disorders [J. Med. Genet. 2003, 40(10):733-40; J. Neuropathol. Exp. Neurol. 2003, 62(9):968-75; J. Neurochem. 2003, 87(2):427-36], autoimmune diseases (N. Engl. J.

10   Med. 2003, 349(16):1526-33; JAMA. 2003, 290(13):1721-8]; gastroesophageal reflux disease [Dig. Dis. Sci. 2003, 48(9):1832-8], cardiovascular diseases [J. Vasc. Surg. 2003, 38(4):827-32], cancerous diseases [Oncogene. 2003, 22(43):6699-703; Br. J. Haematol. 2003, 123(2):288-96], respiratory diseases [Circulation. 2003, 108(15):1839-44], and ophtalmic diseases [Ophthalmology. 2003, 110(10):2040-4;

15   Am. J. Ophthalmol. 2003, 136(4):729-32].

ATPases:

The term "ATPases" refers to enzymes that catalyze the hydrolysis of ATP to ADP, releasing energy that is used in the cell. This group include enzymes such as plasma membrane cation-transporting ATPase, ATP-binding cassette (ABC)

20   transporter, magnesium-ATPase, hydrogen-/sodium-translocating ATPase or ATPase translocating any other elements, arsenite-transporting ATPase, protein-transporting ATPase, DNA translocase, P-type ATPase, and hydrolase, acting on acid anhydrides involved in cellular and subcellular movement.

Pharmaceutical compositions including such proteins or protein encoding

25   sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are associated with impaired conversion of the hydrolysis of ATP to ADP or resulting energy use. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

30   Examples of such diseases include, but are not limited to, infectious diseases such as *helicobacter pylori* ulcers [BMC Gastroenterol. 2003, Nov 6], Neurological, muscular and psychiatric diseases [Int. J. Neurosci. 2003, 13(12):1705-1717; Int. J.

104

Neurosci. 2003, 113(11):1579-1591; Ann. Neurol. 2003, 54(4):494-500], Amyotrophic Lateral Sclerosis [Other Motor Neuron Disord. 2003 4(2):96-9], cardiovascular diseases [J. Nippon. Med. Sch. 2003, 70(5):384-92; Endocrinology. 2003, 144(10):4478-83], metabolic diseases [Mol. Pathol. 2003, 56(5):302-4; Neurosci. Lett.

5   2003, 350(2):105-8], and peptic ulcer disease treated with inhibitors of the gastric H^+-K^+ ATPase (e.g. Omeprazole) responsible for acid secretion in the gastric mucosa.

Carboxylic ester hydrolases:

The phrase carboxylic ester hydrolases" refers to hydrolytic enzymes acting on carboxylic ester bonds such as N-acetylglucosaminylphosphatidylinositol deacetylase,

10   2-acetyl-1-alkylglycerophosphocholine   esterase,   aminoacyl-tRNA   hydrolase, arylesterase, carboxylesterase,   cholinesterase, gluconolactonase, sterol esterase, acetylesterase, carboxymethylenebutenolidase, protein-glutamate methylesterase, lipase, and 6-phosphogluconolactonase.

Pharmaceutical compositions including such proteins or protein encoding

15   sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the hydrolytic cleavage of a covalent bond with accompanying addition of water (-H being added to one product of the cleavage and –OH to the other) is abnormal so that a beneficial effect may be achieved by modulation of such reaction. Antibodies and polynucleotides such as PCR

20   primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, autoimmune neuromuscular disease Myasthenia Gravis, treated with cholinesterase inhibitors.

Hydrolase, acting on ester bonds:

25   The phrase "hydrolase, acting on ester bonds" refers to hydrolytic enzymes acting on ester bonds such as nucleases, sulfuric ester hydrolase, carboxylic ester hydrolases, thiolester hydrolase, phosphoric monoester hydrolase, phosphoric diester hydrolase, triphosphoric monoester hydrolase, diphosphoric monoester hydrolase, and phosphoric triester hydrolase.

30   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the hydrolytic

105

cleavage of a covalent bond with accompanying addition of water (-H being added to one product of the cleavage and -OH to the other), is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Hydrolases:

The term "hydrolases" refers to hydrolytic enzymes such as GPI-anchor transamidase, peptidases, hydrolases, acting on ester bonds, glycosyl bonds, ether bonds, carbon-nitrogen (but not peptide) bonds, acid anhydrides, acid carbon-carbon bonds, acid halide bonds, acid phosphorus-nitrogen bonds, acid sulfur-nitrogen bonds, acid carbon-phosphorus bonds, acid sulfur-sulfur bonds.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the hydrolytic cleavage of a covalent bond with accompanying addition of water (-H being added to one product of the cleavage and -OH to the other) is abnormal. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, cancerous diseases [Cancer. 2003, 98(9):1842-8; Cancer. 2003, 98(9):1822-9], neurological diseases such as Parkinson diseases [J. Neurol. 2003, 250 Suppl 3:III15-III24; J. Neurol. 2003, 250 Suppl 3:III2-III10], endocrinological diseases such as pancreatitis [Pancreas. 2003, 27(4):291-6] or childhood genetic diseases [Eur. J. Pediatr. 1997, 156(12):935-8], coagulation diseases [BMJ. 2003, 327(7421):974-7], cardiovascular diseases [Ann. Intern. Med. 2003, Oct 139(8):670-82], autoimmunity diseases [J. Med. Genet. 2003, 40(10):761-6], and metabolic diseases [Am. J. Hum. Genet. 2001, 69(5):1002-12].

Enzymes:

The term "enzymes' refers to naturally occurring or synthetic macromolecular substance composed mostly of protein, that catalyzes, to various degree of specificity, at least one (bio)chemical reactions at relatively low temperatures. The action of RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic. Nevertheless, enzymes are mainly proteinaceous and are often easily inactivated by

106

heating or by protein-denaturing agents. The substances upon which they act are known as substrates, for which the enzyme possesses a specific binding or active site.

The group of enzymes include various proteins possessing enzymatic activities such as mannosylphosphate transferase, para-hydroxybenzoate:polyprenyltransferase, rieske iron-sulfur protein, imidazoleglycerol-phosphate synthase, sphingosine hydroxylase, tRNA 2'-phosphotransferase, sterol C-24(28) reductase, C-8 sterol isomerase, C-22 sterol desaturase, C-14 sterol reductase, C-3 sterol dehydrogenase (C-4 sterol decarboxylase), 3-keto sterol reductase, C-4 methyl sterol oxidase, dihydronicotinamide riboside quinone reductase, glutamate phosphate reductase, DNA repair enzyme, telomerase, α-ketoacid dehydrogenase, β-alanyl-dopamine synthase, RNA editase, aldo-keto reductase, alkylbase DNA glycosidase, glycogen debranching enzyme, dihydropterin deaminase, dihydropterin oxidase, dimethylnitrosamine demethylase, ecdysteroid UDP-glucosyl/UDP glucuronosyl transferase, glycine cleavage system, helicase, histone deacetylase, mevaldate reductase, monooxygenase, poly(ADP-ribose) glycohydrolase, pyruvate dehydrogenase, serine esterase, sterol carrier protein X-related thiolase, transposase, tyramine-β hydroxylase, para-aminobenzoic acid (PABA) synthase, glu-tRNA(gln) amidotransferase, molybdopterin cofactor sulfurase, lanosterol 14-α-demethylase, aromatase, 4-hydroxybenzoate octaprenyltransferase, 7,8-dihydro-8-oxoguanine-triphosphatase, CDP-alcohol phosphotransferase, 2,5-diamino-6-(ribosylamino)-4(3H)-pyrimidonone 5'-phosphate deaminase, diphosphoinositol polyphosphate phosphohydrolase, γ-glutamyl carboxylase, small protein conjugating enzyme, small protein activating enzyme, 1-deoxyxylulose-5-phosphate synthase, 2'-phosphotransferase, 2-octoprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, 3,4 dihydroxy-2-butanone-4-phosphate synthase, 4-amino-4-deoxychorismate lyase, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase, ADP-L-glycero-D-manno-heptose synthase, D-erythro-7,8-dihydroneopterin triphosphate 2'-epimerase, N-ethylmaleimide reductase, O-antigen ligase, O-antigen polymerase, UDP-2,3-diacylglucosamine hydrolase, arsenate reductase, carnitine racemase, cobalamin [5'-phosphate] synthase, cobinamide phosphate guanylyltransferase, enterobactin synthetase, enterochelin esterase, enterochelin synthetase, glycolate oxidase, integrase, lauroyl transferase, peptidoglycan synthetase,

phosphopantetheinyltransferase, phosphoglucosamine mutase, phosphoheptose isomerase, quinolinate synthase, siroheme synthase, N-acylmannosamine-6-phosphate 2-epimerase, N-acetyl-anhydromuramoyl-L-alanine amidase, carbon-phosphorous lyase, heme-copper terminal oxidase, disulfide oxidoreductase, phthalate dioxygenase reductase, sphingosine-1-phosphate lyase, molybdopterin oxidoreductase, dehydrogenase, NADPH oxidase, naringenin-chalcone synthase, N-ethylammeline chlorohydrolase, polyketide synthase, aldolase, kinase, phosphatase, CoA-ligase, oxidoreductase, transferase, hydrolase, lyase, isomerase, ligase, ATPase, sulfhydryl oxidase, lipoate-protein ligase, δ-1-pyrroline-5-carboxyate synthetase, lipoic acid synthase, and tRNA dihydrouridine synthase.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which can be ameliorated by modulating the activity of various enzymes which are involved both in enzymatic processes inside cells as well as in cell signaling. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Cytoskeletal proteins:

The term "cytoskeletal proteins" refers to proteins involved in the structure formation of the cytoskeleton.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are caused or due to abnormalities in cytoskeleton, including cancerous cells, and diseased cells such as cells that do not propagate, grow or function normally. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, liver diseases such as cholestatic diseases [Lancet. 2003, 362(9390):1112-9], vascular diseases [J. Cell Biol. 2003, 162(6):1111-22], endocrinological diseases [Cancer Res. 2003, 63(16):4836-41], neuromuscular disorders such as muscular dystrophy [Neuromuscul. Disord. 2003, 13(7-8):579-88], or myopathy [Neuromuscul. Disord. 2003, 13(6):456-67] neurological

108

disorders such as Alzheimer's disease [J. Alzheimers Dis. 2003, 5(3):209-28], cardiac disorders [J. Am. Coll. Cardiol. 2003, 42(2):319-27], skin disorders [J. Am. Coll. Cardiol. 2003, 42(2):319-27], and cancer [Proteomics. 2003, 3(6):979-90].

Structural proteins:

The term "structural proteins" refers to proteins involved in the structure formation of the cell, such as structural proteins of ribosome, cell wall structural proteins, structural proteins of cytoskeleton, extracellular matrix structural proteins, extracellular matrix glycoproteins, amyloid proteins, plasma proteins, structural proteins of eye lens, structural protein of chorion (sensu Insecta), structural protein of cuticle (sensu Insecta), puparial glue protein (sensu Diptera), structural proteins of bone, yolk proteins, structural proteins of muscle, structural protein of vitelline membrane (sensu Insecta), structural proteins of peritrophic membrane (sensu Insecta), and structural proteins of nuclear pores.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases which are caused by abnormalities in cytoskeleton, including cancerous cells, and diseased cells such as cells that do not propagate, grow or function normally. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, blood vessels diseases such as aneurysms [Cardiovasc. Res. 2003, 60(1):205-13], joint diseases [Rheum. Dis. Clin. North Am. 2003, 29(3):631-45], muscular diseases such as muscular dystrophies [Curr. Opin. Clin. Nutr. Metab. Care. 2003, 6(4):435-9], neuronal diseases such as encephalitis [Neurovirol. 2003, 9(2):274-83], retinitis pigmentosa [Dev. Ophthalmol. 2003, 37:109-25], and infectious diseases [J. Virol. Methods. 2003, 109(1):75-83; FEMS Immunol. Med. Microbiol. 2003, 35(2):125-30; J. Exp. Med. 2003, 197(5):633-42].

Ligands:

The term "ligands" refers to proteins that bind to another chemical entity to form a larger complex, involved in various biological processes, such as signal transduction, metabolism, growth and differentiation, etc. This group of proteins includes opioid peptides, baboon receptor ligand, branchless receptor ligand,

109

breathless receptor ligand, ephrin, frizzled receptor ligand, frizzled-2 receptor ligand, heartless receptor ligand, Notch receptor ligand, patched receptor ligand, punt receptor ligand, Ror receptor ligand, saxophone receptor ligand, SE20 receptor ligand, sevenless receptor ligand, smooth receptor ligand, thickveins receptor ligand, Toll

5   receptor ligand, Torso receptor ligand, death receptor ligand, scavenger receptor ligand, neuroligin, integrin ligand, hormones, pheromones, growth factors, and sulfonylurea receptor ligand.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering

10   expression of such proteins, may be used to treat diseases involved in impaired hormone function or diseases which involve abnormal secretion of proteins which may be due to abnormal presence, absence or impaired normal response to normal levels of secreted proteins. Those secreted proteins include hormones, neurotransmitters, and various other proteins secreted by cells to the extracellular environment. Antibodies and

15   polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, analgesia inhibited by orphanin FQ/nociceptin [Shane R., et al., (2001) Brain Res., 907(1-2):109-16], stroke protected by estrogen [Alkayed N. J., et al., (2001) J. Neurosci., 21(19):7543-50],

20   atherosclerosis associated with growth hormone deficiency [Elhadd T .A., et al., (2001) J. Clin. Endocrinol. Metab., 86(9):4223-32], diabetes inhibited by α-galactosylceramide [Hong S., et al., (2001) Nat. Med., 7(9):1052-6], and Huntington's disease associated with huntingtin deficiency [Rao D. S., et al., (2001) Mol. Cell Biol., 21(22):7796-806].

Signal transducer:

25   The term "signal transducers" refers to proteins such as activin inhibitors, receptor-associated proteins, α-2 macroglobulin receptors, morphogens, quorum sensing signal generators, quorum sensing response regulators, receptor signaling proteins, ligands, receptors, two-component sensor molecules, and two-component response regulators.

30   Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases in which the signal-transduction

110

is impaired, either as a cause, or as a result of the disease. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, altered sexual dimorphism associated with signal transducer and activator of transcription 5b [Udy G. B., et al., (1997) Proc. Natl. Acad. Sci. U S A, 94(14):7239-44], multiple sclerosis associated with sgp130 deficiency [Padberg F., et al., (1999) J. Neuroimmunol., 99(2):218-23], intestinal inflammation associated with elevated signal transducer and activator of transcription 3 activity [Suzuki A., et al., (2001) J Exp Med, 193(4):471-81], carcinoid tumor inhibited by increased signal transducer and activators of transcription 1 and 2 [Zhou Y., et al., (2001) Oncology, 60(4):330-8], and esophageal cancer associated with loss of EGF-STAT1 pathway [Watanabe G., et al., (2001) Cancer J., 7(2):132-9].

RNA polymerase II transcription factors:

The phrase "RNA polymerase II transcription factors" refers to proteins such as specific and non-specific RNA polymerase II transcription factors, enhancer binding, ligand-regulated transcription factor, and general RNA polymerase II transcription factors.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving impaired function of RNA polymerase II transcription factors. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, cardiac diseases [Cell Cycle. 2003, 2(2):99-104], xeroderma pigmentosum [Bioessays. 2001, 23(8):671-3; Biochim. Biophys. Acta. 1997, 1354(3):241-51], muscular atrophy [J. Cell Biol. 2001, 152(1):75-85], neurological diseases such as Alzheimer's disease [Front Biosci. 2000, 5:D244-57], cancerous diseases such as breast cancer [Biol. Chem. 1999, 380(2):117-28], and autoimmune disorders [Clin. Exp. Immunol. 1997, 109(3):488-94].

RNA binding proteins:

The phrase "RNA binding proteins" refers to RNA binding proteins involved in splicing and translation regulation such as tRNA binding proteins, RNA helicases,

111

double-stranded RNA and single-stranded RNA binding proteins, mRNA binding proteins, snRNA cap binding proteins, 5S RNA and 7S RNA binding proteins, poly-pyrimidine tract binding proteins, snRNA binding proteins, and AU-specific RNA binding proteins.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving transcription and translation factors such as helicases, isomerases, histones and nucleases, diseases where there is impaired transcription, splicing, post-transcriptional processing, translation or stability of the RNA. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such diseases include, but are not limited to, cancerous diseases such as lymphomas [Tumori. 2003, 89(3):278-84], prostate cancer [Prostate. 2003, 57(1):80-92] or lung cancer [J. Pathol. 2003, 200(5):640-6], blood diseases, such as fanconi anemia [Curr. Hematol. Rep. 2003, 2(4):335-40], cardiovascular diseases such as atherosclerosis [J. Thromb. Haemost. 2003, 1(7):1381-90] muscle diseases [Trends Cardiovasc. Med. 2003, 13(5):188-95] and brain and neuronal diseases [Trends Cardiovasc. Med. 2003, 13(5):188-95; Neurosci. Lett. 2003, 342(1-2):41-4].

Nucleic acid binding proteins:

The phrase "nucleic acid binding proteins" refers to proteins involved in RNA and DNA synthesis and expression regulation such as transcription factors, RNA and DNA binding proteins, zinc fingers, helicase, isomerase, histones, nucleases, ribonucleoproteins, and transcription and translation factors.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving DNA or RNA binding proteins such as: helicases, isomerases, histones and nucleases, for example diseases where there is abnormal replication or transcription of DNA and RNA respectively. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

112

Examples of such diseases include, but are not limited to, neurological diseases such as renitis pigmentoas [Am. J. Ophthalmol. 2003, 136(4):678-87] parkinsonism [Proc. Natl. Acad. Sci. U S A. 2003, 100(18):10347-52], Alzheimer [J. Neurosci. 2003, 23(17):6914-27] and canavan diseases [Brain Res Bull. 2003, 61(4):427-35], cancerous diseases such as leukemia [Anticancer Res. 2003, 23(4):3419-26] or lung cancer [J. Pathol. 2003, 200(5):640-6], miopathy [Neuromuscul Disord. 2003, 13(7-8):559-67] and liver diseases [J. Pathol. 2003, 200(5):553-60].

Proteins involved in Metabolism:

The phrase "proteins involved in metabolism" refers to proteins involved in the totality of the chemical reactions and physical changes that occur in living organisms, comprising anabolism and catabolism; may be qualified to mean the chemical reactions and physical processes undergone by a particular substance, or class of substances, in a living organism. This group includes proteins involved in the reactions of cell growth and maintenance such as: metabolism resulting in cell growth, carbohydrate metabolism, energy pathways, electron transport, nucleobase, nucleoside, nucleotide and nucleic acid metabolism, protein metabolism and modification, amino acid and derivative metabolism, protein targeting, lipid metabolism, aromatic compound metabolism, one-carbon compound metabolism, coenzymes and prosthetic group metabolism, sulfur metabolism, phosphorus metabolism, phosphate metabolism, oxygen and radical metabolism, xenobiotic metabolism, nitrogen metabolism, fat body metabolism (sensu Insecta), protein localization, catabolism, biosynthesis, toxin metabolism , methylglyoxal metabolism, cyanate metabolism, glycolate metabolism, carbon utilization and antibiotic metabolism.

Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat diseases involving cell metabolism. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases.

Examples of such metabolism-related diseases include, but are not limited to, multisystem mitochondrial disorder caused by mitochondrial DNA cytochrome C

113

oxidase II deficiency [Campos Y., et al., (2001) Ann. Neurol. 50(3):409-13], conduction defects and ventricular dysfunction in the heart associated with heterogeneous connexin43 expression [Gutstein D. E., et al., (2001) Circulation, 104(10):1194-9], atherosclerosis associated with growth suppressor p27 deficiency
5   [Diez-Juan A., and Andres V. (2001) FASEB J., 15(11):1989-95], colitis associated with glutathione peroxidase deficiency [Esworthy R. S., et al., (2001) Am. J. Physiol. Gastrointest. Liver Physiol., 281(3):G848-55], systemic lupus erythematosus associated with deoxyribonuclease I deficiency [Yasutomo K., et al., (2001) Nat. Genet., 28(4):313-4], alcoholic pancreatitis [Pancreas. 2003, 27(4):281-5],
10 amyloidosis and diseases that are related to amyloid metabolism, such as FMF, atherosclerosis, diabetes, and especially diabetes long term consequences, neurological diseases such as Creutzfeldt-Jakob disease, and Parkinson or Rasmussen's encephalitis.

Cell growth and/or maintenance proteins:

15   The phrase "Cell growth and/or maintenance proteins" refers to proteins involved in any biological process required for cell survival, growth and maintenance, including proteins involved in biological processes such as cell organization and biogenesis, cell growth, cell proliferation, metabolism, cell cycle, budding, cell shape and cell size control, sporulation (sensu Saccharomyces), transport, ion homeostasis,
20 autophagy, cell motility, chemi-mechanical coupling, membrane fusion, cell-cell fusion, and stress response.

        Pharmaceutical compositions including such proteins or protein encoding sequences, antibodies directed against such proteins or polynucleotides capable of altering expression of such proteins, may be used to treat or prevent diseases such as cancer,
25 degenerative diseases, for example neurodegenerative diseases or conditions associated with aging, or alternatively, diseases wherein apoptosis which should have taken place, does not take place. Antibodies and polynucleotides such as PCR primers and molecular probes designed to identify such proteins or protein encoding sequences may be used for diagnosis of such diseases, detection of pre-disposition to a disease, and determination of
30 the stage of a disease.

        Examples of such diseases include, but are not limited to, ataxia-telangiectasia associated with ataxia-telangiectasia mutated deficiency [Hande et al., (2001) Hum.

114

Mol. Genet., 10(5):519-28], osteoporosis associated with osteonectin deficiency [Delany et al., (2000) J. Clin. Invest., 105(7):915-23], arthritis caused by membrane-bound matrix metalloproteinase deficiency [Holmbeck et al., (1999) Cell, 99(1):81-92], defective stratum corneum and early neonatal death associated with transglutaminase 1 deficiency [Matsuki et al., (1998) Proc. Natl. Acad. Sci. U S A, 95(3):1044-9], and Alzheimer's disease associated with estrogen [Simpkins et al., (1997) Am. J. Med., 103(3A):19S-25S].

Chaperones

Information derived from proteins such as ribosomal chaperone, peptidylprolyl isomerase, lectin-binding chaperone, nucleosome assembly chaperone, chaperonin ATPase, cochaperone, heat shock protein, HSP70/HSP90 organizing protein, fimbrial chaperone, metallochaperone, tubulin folding, HSC70-interacting protein can be used to diagnose/treat diseases involving pathological conditions, which are associated with non-normal protein activity or structure. Binding of the products of the proteins of this family, or antibodies reactive therewith, can modulate a plurality of protein activities as well as change protein structure. Alternatively, diseases in which there is abnormal degradation of other proteins, which may cause non-normal accumulation of various proteinaceous products in cells, caused non- normal (prolonged or shortened) activity of proteins, etc.

Example of diseases that involve chaperones are cancerous diseases, such as prostate cancer (Semin Oncol. 2003 Oct;30(5):709-16.); infectious diseases, such as prion infection (EMBO J. 2003 Oct 15;22(20):5435-5445.); neurological syndromes (J Neuropathol Exp Neurol. 2003 Jul;62(7):751-64.; Antioxid Redox Signal. 2003 Jun;5(3):337-48.; J Neurochem. 2003 Jul;86(2):394-404.)

Variants of proteins which accumulate an element/compound

Variant proteins which their wild type version naturally binds a certain compound or element inside the cell for storage of accumulation may have terapoetic effect as secreted variants. Ferritin, accumulates iron inside the cells. A secreted variant of this protein is expected to bind plasma iron, reduce its levels and therefore have a desired therapeutic effect in the syndrome of Hemosiderosis characterized by high levels of iron in the blood.

115

Diseases that may be treated/diagnosed using the biomolecular sequences of the

present invention

Inflammatory diseases

Examples of inflammatory diseases include, but are not limited to, chronic inflammatory diseases and acute inflammatory diseases.

Inflammatory diseases associated with hypersensitivity

Examples of hypersensitivity include, but are not limited to, Types I-IV hypersensitivity, immediate hypersensitivity, antibody mediated hypersensitivity, immune complex mediated hypersensitivity, T lymphocyte mediated hypersensitivity and DTH. An example of type I or immediate hypersensitivity is asthma. Examples of type II hypersensitivity include, but are not limited to, rheumatoid diseases, rheumatoid autoimmune diseases, rheumatoid arthritis [Krenn V. *et al.*, Histol Histopathol 2000 Jul;15 (3):791], spondylitis, ankylosing spondylitis [Jan Voswinkel *et al.*, Arthritis Res 2001; 3 (3): 189], systemic diseases, systemic autoimmune diseases, systemic lupus erythematosus [Erikson J. *et al.*, Immunol Res 1998;17 (1-2):49], sclerosis, systemic sclerosis [Renaudineau Y. *et al.*, Clin Diagn Lab Immunol. 1999 Mar;6 (2):156; Chan OT. *et al.*, Immunol Rev 1999 Jun;169:107], glandular diseases, glandular autoimmune diseases, pancreatic autoimmune diseases, diabetes, Type I diabetes [Zimmet P. Diabetes Res Clin Pract 1996 Oct;34 Suppl:S125], thyroid diseases, autoimmune thyroid diseases, Graves' disease [Orgiazzi J. Endocrinol Metab Clin North Am 2000 Jun;29 (2):339], thyroiditis, spontaneous autoimmune thyroiditis [Braley-Mullen H. and Yu S, J Immunol 2000 Dec 15;165 (12):7262], Hashimoto's thyroiditis [Toyoda N. *et al.*, Nippon Rinsho 1999 Aug;57 (8):1810], myxedema, idiopathic myxedema [Mitsuma T. Nippon Rinsho. 1999 Aug;57 (8):1759], autoimmune reproductive diseases, ovarian diseases, ovarian autoimmunity [Garza KM. *et al.*, J Reprod Immunol 1998 Feb;37 (2):87], autoimmune anti-sperm infertility [Diekman AB. *et al.*, Am J Reprod Immunol. 2000 Mar;43 (3):134], repeated fetal loss [Tincani A. *et al.*, Lupus 1998;7 Suppl 2:S107-9], neurodegenerative diseases, neurological diseases, neurological autoimmune diseases, multiple sclerosis [Cross AH. *et al.*, J Neuroimmunol 2001 Jan 1;112 (1-2):1], Alzheimer's disease [Oron L. *et al.*, J Neural Transm Suppl. 1997;49:77], myasthenia gravis [Infante AJ. And Kraig E, Int Rev Immunol 1999;18 (1-2):83], motor

116

neuropathies [Kornberg AJ. J Clin Neurosci. 2000 May;7 (3):191], Guillain-Barre syndrome, neuropathies and autoimmune neuropathies [Kusunoki S. Am J Med Sci. 2000 Apr;319 (4):234], myasthenic diseases, Lambert-Eaton myasthenic syndrome [Takamori M. Am J Med Sci. 2000 Apr;319 (4):204], paraneoplastic neurological

5   diseases, cerebellar atrophy, paraneoplastic cerebellar atrophy, non-paraneoplastic stiff man syndrome, cerebellar atrophies, progressive cerebellar atrophies, encephalitis, Rasmussen's encephalitis, amyotrophic lateral sclerosis, Sydeham chorea, Gilles de la Tourette syndrome, polyendocrinopathies, autoimmune polyendocrinopathies [Antoine JC. and Honnorat J. Rev Neurol (Paris) 2000 Jan;156

10   (1):23], neuropathies, dysimmune neuropathies [Nobile-Orazio E. *et al.,* Electroencephalogr Clin Neurophysiol Suppl 1999;50:419], neuromyotonia, acquired neuromyotonia, arthrogryposis multiplex congenita [Vincent A. *et al.,* Ann N Y Acad Sci. 1998 May 13;841:482], cardiovascular diseases, cardiovascular autoimmune diseases, atherosclerosis [Matsuura E. *et al.,* Lupus. 1998;7 Suppl 2:S135],

15   myocardial infarction [Vaarala O. Lupus. 1998;7 Suppl 2:S132], thrombosis [Tincani A. *et al.,* Lupus 1998;7 Suppl 2:S107-9], granulomatosis, Wegener's granulomatosis, arteritis, Takayasu's arteritis and Kawasaki syndrome [Praprotnik S. *et al.,* Wien Klin Wochenschr 2000 Aug 25;112 (15-16):660], anti-factor VIII autoimmune disease [Lacroix-Desmazes S. *et al.,* Semin Thromb Hemost.2000;26 (2):157], vasculitises,

20   necrotizing small vessel vasculitises, microscopic polyangiitis, Churg and Strauss syndrome, glomerulonephritis, pauci-immune focal necrotizing glomerulonephritis, crescentic glomerulonephritis [Noel LH. Ann Med Interne (Paris). 2000 May;151 (3):178], antiphospholipid syndrome [Flamholz R. *et al.,* J Clin Apheresis 1999;14 (4):171], heart failure, agonist-like β-adrenoceptor antibodies in heart failure

25   [Wallukat G. *et al.,* Am J Cardiol. 1999 Jun 17;83 (12A):75H], thrombocytopenic purpura [Moccia F. Ann Ital Med Int. 1999 Apr-Jun;14 (2):114], hemolytic anemia, autoimmune hemolytic anemia [Efremov DG. *et al.,* Leuk Lymphoma 1998 Jan;28 (3-4):285], gastrointestinal diseases, autoimmune diseases of the gastrointestinal tract, intestinal diseases, chronic inflammatory intestinal disease [Garcia Herola A. *et al.,*

30   Gastroenterol Hepatol. 2000 Jan;23 (1):16], celiac disease [Landau YE. and Shoenfeld Y. Harefuah 2000 Jan 16;138 (2):122], autoimmune diseases of the musculature, myositis, autoimmune myositis, Sjogren's syndrome [Feist E. *et al.,* Int

117

Arch Allergy Immunol 2000 Sep;123 (1):92], smooth muscle autoimmune disease [Zauli D. *et al.*, Biomed Pharmacother 1999 Jun;53 (5-6):234], hepatic diseases, hepatic autoimmune diseases, autoimmune hepatitis [Manns MP. J Hepatol 2000 Aug;33 (2):326] and primary biliary cirrhosis [Strassburg CP. *et al.*, Eur J Gastroenterol Hepatol. 1999 Jun;11 (6):595].

Examples of type IV or T cell mediated hypersensitivity, include, but are not limited to, rheumatoid diseases, rheumatoid arthritis [Tisch R, McDevitt HO. Proc Natl Acad Sci U S A 1994 Jan 18;91 (2):437], systemic diseases, systemic autoimmune diseases, systemic lupus erythematosus [Datta SK., Lupus 1998;7 (9):591], glandular diseases, glandular autoimmune diseases, pancreatic diseases, pancreatic autoimmune diseases, Type 1 diabetes [Castano L. and Eisenbarth GS. Ann. Rev. Immunol. 8:647], thyroid diseases, autoimmune thyroid diseases, Graves' disease [Sakata S. *et al.*, Mol Cell Endocrinol 1993 Mar;92 (1):77], ovarian diseases [Garza KM. *et al.*, J Reprod Immunol 1998 Feb;37 (2):87], prostatitis, autoimmune prostatitis [Alexander RB. *et al.*, Urology 1997 Dec;50 (6):893], polyglandular syndrome, autoimmune polyglandular syndrome, Type I autoimmune polyglandular syndrome [Hara T. *et al.*, Blood. 1991 Mar 1;77 (5):1127], neurological diseases, autoimmune neurological diseases, multiple sclerosis, neuritis, optic neuritis [Soderstrom M. *et al.*, J Neurol Neurosurg Psychiatry 1994 May;57 (5):544], myasthenia gravis [Oshima M. *et al.*, Eur J Immunol 1990 Dec;20 (12):2563], stiff-man syndrome [Hiemstra HS. *et al.*, Proc Natl Acad Sci U S A 2001 Mar 27;98 (7):3988], cardiovascular diseases, cardiac autoimmunity in Chagas' disease [Cunha-Neto E. *et al.*, J Clin Invest 1996 Oct 15;98 (8):1709], autoimmune thrombocytopenic purpura [Semple JW. *et al.*, Blood 1996 May 15;87 (10):4245], anti-helper T lymphocyte autoimmunity [Caporossi AP. *et al.*, Viral Immunol 1998;11 (1):9], hemolytic anemia [Sallah S. *et al.*, Ann Hematol 1997 Mar;74 (3):139], hepatic diseases, hepatic autoimmune diseases, hepatitis, chronic active hepatitis [Franco A. *et al.*, Clin Immunol Immunopathol 1990 Mar;54 (3):382], biliary cirrhosis, primary biliary cirrhosis [Jones DE. Clin Sci (Colch) 1996 Nov;91 (5):551], nephric diseases, nephric autoimmune diseases, nephritis, interstitial nephritis [Kelly CJ. J Am Soc Nephrol 1990 Aug;1 (2):140], connective tissue diseases, ear diseases, autoimmune connective tissue diseases, autoimmune ear disease [Yoo TJ. *et al.*, Cell Immunol

118

1994 Aug;157 (1):249], disease of the inner ear [Gloddek B. *et al.*, Ann N Y Acad Sci 1997 Dec 29;830:266], skin diseases, cutaneous diseases, dermal diseases, bullous skin diseases, pemphigus vulgaris, bullous pemphigoid and pemphigus foliaceus.

Examples of delayed type hypersensitivity include, but are not limited to, contact dermatitis and drug eruption.

Autoimmune diseases

Examples of autoimmune diseases include, but are not limited to, cardiovascular diseases, rheumatoid diseases, glandular diseases, gastrointestinal diseases, cutaneous diseases, hepatic diseases, neurological diseases, muscular diseases, nephric diseases, diseases related to reproduction, connective tissue diseases and systemic diseases.

Examples of autoimmune cardiovascular and blood diseases include, but are not limited to atherosclerosis [Matsuura E. *et al.*, Lupus. 1998;7 Suppl 2:S135], myocardial infarction [Vaarala O. Lupus. 1998;7 Suppl 2:S132], thrombosis [Tincani A. *et al.*, Lupus 1998;7 Suppl 2:S107-9], Wegener's granulomatosis, Takayasu's arteritis, Kawasaki syndrome [Praprotnik S. *et al.*, Wien Klin Wochenschr 2000 Aug 25;112 (15-16):660], anti-factor VIII autoimmune disease [Lacroix-Desmazes S. *et al.*, Semin Thromb Hemost.2000;26 (2):157], necrotizing small vessel vasculitis, microscopic polyangiitis, Churg and Strauss syndrome, pauci-immune focal necrotizing and crescentic glomerulonephritis [Noel LH. Ann Med Interne (Paris). 2000 May;151 (3):178], antiphospholipid syndrome [Flamholz R. *et al.*, J Clin Apheresis 1999;14 (4):171], antibody-induced heart failure [Wallukat G. *et al.*, Am J Cardiol. 1999 Jun 17;83 (12A):75H], thrombocytopenic purpura [Moccia F. Ann Ital Med Int. 1999 Apr-Jun;14 (2):114; Semple JW. *et al.*, Blood 1996 May 15;87 (10):4245], autoimmune hemolytic anemia [Efremov DG. *et al.*, Leuk Lymphoma 1998 Jan;28 (3-4):285; Sallah S. *et al.*, Ann Hematol 1997 Mar;74 (3):139], cardiac autoimmunity in Chagas' disease [Cunha-Neto E. *et al.*, J Clin Invest 1996 Oct 15;98 (8):1709) and anti-helper T lymphocyte autoimmunity [Caporossi AP. *et al.*, Viral Immunol 1998;11 (1):9].

Examples of autoimmune rheumatoid diseases include, but are not limited to rheumatoid arthritis [Krenn V. *et al.*, Histol Histopathol 2000 Jul;15 (3):791; Tisch R,

119

McDevitt HO. Proc Natl Acad Sci units S A 1994 Jan 18;91 (2):437) and ankylosing spondylitis [Jan Voswinkel *et al.*, Arthritis Res 2001; 3 (3): 189].

Examples of autoimmune glandular diseases include, but are not limited to, pancreatic disease, Type I diabetes, Type II diabetes, thyroid disease, Graves' disease, thyroiditis, spontaneous autoimmune thyroiditis, Hashimoto's thyroiditis, idiopathic myxedema, ovarian autoimmunity, autoimmune anti-sperm infertility, autoimmune prostatitis and Type I autoimmune polyglandular syndrome. diseases include, but are not limited to autoimmune diseases of the pancreas, Type 1 diabetes [Castano L. and Eisenbarth GS. Ann. Rev. Immunol. 8:647; Zimmet P. Diabetes Res Clin Pract 1996 Oct;34 Suppl:S125], autoimmune thyroid diseases, Graves' disease [Orgiazzi J. Endocrinol Metab Clin North Am 2000 Jun;29 (2):339; Sakata S. *et al.*, Mol Cell Endocrinol 1993 Mar;92 (1):77], spontaneous autoimmune thyroiditis [Braley-Mullen H. and Yu S, J Immunol 2000 Dec 15;165 (12):7262], Hashimoto's thyroiditis [Toyoda N. *et al.*, Nippon Rinsho 1999 Aug;57 (8):1810], idiopathic myxedema [Mitsuma T. Nippon Rinsho. 1999 Aug;57 (8):1759], ovarian autoimmunity [Garza KM. *et al.*, J Reprod Immunol 1998 Feb;37 (2):87], autoimmune anti-sperm infertility [Diekman AB. *et al.*, Am J Reprod Immunol. 2000 Mar;43 (3):134], autoimmune prostatitis [Alexander RB. *et al.*, Urology 1997 Dec;50 (6):893) and Type I autoimmune polyglandular syndrome [Hara T. *et al.*, Blood. 1991 Mar 1;77 (5):1127].

Examples of autoimmune gastrointestinal diseases include, but are not limited to, chronic inflammatory intestinal diseases [Garcia Herola A. *et al.*, Gastroenterol Hepatol. 2000 Jan;23 (1):16], celiac disease [Landau YE. and Shoenfeld Y. Harefuah 2000 Jan 16;138 (2):122], colitis, ileitis and Crohn's disease and ulcerative colitis.

Examples of autoimmune cutaneous diseases include, but are not limited to, autoimmune bullous skin diseases, such as, but are not limited to, pemphigus vulgaris, bullous pemphigoid and pemphigus foliaceus.

Examples of autoimmune hepatic diseases include, but are not limited to, hepatitis, autoimmune chronic active hepatitis [Franco A. *et al.*, Clin Immunol Immunopathol 1990 Mar;54 (3):382], primary biliary cirrhosis [Jones DE. Clin Sci (Colch) 1996 Nov;91 (5):551; Strassburg CP. *et al.*, Eur J Gastroenterol Hepatol. 1999 Jun;11 (6):595) and autoimmune hepatitis [Manns MP. J Hepatol 2000 Aug;33 (2):326].

120

Examples of autoimmune neurological diseases include, but are not limited to, multiple sclerosis [Cross AH. *et al.*, J Neuroimmunol 2001 Jan 1;112 (1-2):1], Alzheimer's disease [Oron L. *et al.*, J Neural Transm Suppl. 1997;49:77], myasthenia gravis [Infante AJ. And Kraig E, Int Rev Immunol 1999;18 (1-2):83; Oshima M. *et al.*, Eur J Immunol 1990 Dec;20 (12):2563], neuropathies, motor neuropathies [Kornberg AJ. J Clin Neurosci. 2000 May;7 (3):191], Guillain-Barre syndrome and autoimmune neuropathies [Kusunoki S. Am J Med Sci. 2000 Apr;319 (4):234], myasthenia, Lambert-Eaton myasthenic syndrome [Takamori M. Am J Med Sci. 2000 Apr;319 (4):204], paraneoplastic neurological diseases, cerebellar atrophy, paraneoplastic cerebellar atrophy and stiff-man syndrome [Hiemstra HS. *et al.*, Proc Natl Acad Sci units S A 2001 Mar 27;98 (7):3988], non-paraneoplastic stiff man syndrome, progressive cerebellar atrophies, encephalitis, Rasmussen's encephalitis, amyotrophic lateral sclerosis, Sydeham chorea, Gilles de la Tourette syndrome and autoimmune polyendocrinopathies [Antoine JC. and Honnorat J. Rev Neurol (Paris) 2000 Jan;156 (1):23], dysimmune neuropathies [Nobile-Orazio E. *et al.*, Electroencephalogr Clin Neurophysiol Suppl 1999;50:419], acquired neuromyotonia, arthrogryposis multiplex congenita [Vincent A. *et al.*, Ann N Y Acad Sci. 1998 May 13;841:482], neuritis, optic neuritis [Soderstrom M. *et al.*, J Neurol Neurosurg Psychiatry 1994 May;57 (5):544) multiple sclerosis and neurodegenerative diseases.

Examples of autoimmune muscular diseases include, but are not limited to, myositis, autoimmune myositis and primary Sjogren's syndrome [Feist E. *et al.*, Int Arch Allergy Immunol 2000 Sep;123 (1):92) and smooth muscle autoimmune disease [Zauli D. *et al.*, Biomed Pharmacother 1999 Jun;53 (5-6):234].

Examples of autoimmune nephric diseases include, but are not limited to, nephritis and autoimmune interstitial nephritis [Kelly CJ. J Am Soc Nephrol 1990 Aug;1 (2):140], glommerular nephritis.

Examples of autoimmune diseases related to reproduction include, but are not limited to, repeated fetal loss [Tincani A. *et al.*, Lupus 1998;7 Suppl 2:S107-9].

Examples of autoimmune connective tissue diseases include, but are not limited to, ear diseases, autoimmune ear diseases [Yoo TJ. *et al.*, Cell Immunol 1994 Aug;157 (1):249) and autoimmune diseases of the inner ear [Gloddek B. *et al.*, Ann N Y Acad Sci 1997 Dec 29;830:266].

Examples of autoimmune systemic diseases include, but are not limited to, systemic lupus erythematosus [Erikson J. *et al.*, Immunol Res 1998;17 (1-2):49) and systemic sclerosis [Renaudineau Y. *et al.*, Clin Diagn Lab Immunol. 1999 Mar;6 (2):156; Chan OT. *et al.*, Immunol Rev 1999 Jun;169:107].

5   *Infectious diseases*

Examples of infectious diseases include, but are not limited to, chronic infectious diseases, subacute infectious diseases, acute infectious diseases, viral diseases, bacterial diseases, protozoan diseases, parasitic diseases, fungal diseases, mycoplasma diseases, and prion diseases.

10

Graft rejection diseases

Examples of diseases associated with transplantation of a graft include, but are not limited to, graft rejection, chronic graft rejection, subacute graft rejection, hyperacute graft rejection, acute graft rejection, and graft versus host disease.

15   *Allergic diseases*

Examples of allergic diseases include, but are not limited to, asthma, hives, urticaria, pollen allergy, dust mite allergy, venom allergy, cosmetics allergy, latex allergy, chemical allergy, drug allergy, insect bite allergy, animal dander allergy, stinging plant allergy, poison ivy allergy and food allergy.

20   *Cancerous diseases*

Examples of cancer include but are not limited to carcinoma, lymphoma, blastoma, sarcoma, and leukemia. Particular examples of cancerous diseases but are not limited to: Myeloid leukemia such as Chronic myelogenous leukemia. Acute myelogenous leukemia with maturation. Acute promyelocytic leukemia, Acute

25   nonlymphocytic leukemia with increased basophils, Acute monocytic leukemia. Acute myelomonocytic leukemia with eosinophilia; malignant lymphoma, such as Birkitt's Non-Hodgkin's; Lymphoctyic leukemia, such as acute lumphoblastic leukemia. Chronic lymphocytic leukemia; Myeloproliferative diseases, such as Solid tumors Benign Meningioma, Mixed tumors of salivary gland, Colonic adenomas;

30   Adenocarcinomas, such as Small cell lung cancer, Kidney, Uterus, Prostate, Bladder, Ovary, Colon, Sarcomas, Liposarcoma, myxoid, Synovial sarcoma, Rhabdomyosarcoma (alveolar), Extraskeletel myxoid chonodrosarcoma, Ewing's

122

tumor; other include Testicular and ovarian dysgerminoma, Retinoblastoma, Wilms'
tumor, Neuroblastoma, Malignant melanoma, Mesothelioma, breast, skin, prostate,
and ovarian.

EXAMPLE 9

Microarray analysis based validation of the antisense dataset

A microarray-based analysis using oligonucleotide probes that hybridize to the
target in a strand-specific manner, was conducted in order to experimentally validate
the predicted antisense/sense pairs of the database. Two complementary 60-mer
oligonucleotide probes derived from the predicted overlap region of the
sense/antisense pairs, were designed. Single 60-mer oligonucleotides were previously
shown to offer reliability and sensitivity for detecting specific transcripts (T. R.
Hughes, *et al.*, *Nature Biotech.* **19**, 342 (2001).) Initially only pairs of clusters with
an overlap greater than 60 bases (2,464 pairs agree with this restriction) were selected
for array construction. The overlap region of each antisense pair was then verified for
the presence of 60-mer oligonucleotides that matched a set of standards, such as
minimal sequence similarity elsewhere in the human genome, uniform GC-content
and Tm, and absence of palindromic sequences, in order to maximize the
hybridization specificity. Oligonucleotide probes meeting the criteria set forth were
identified for 1,211 sense/antisense pairs and a random sample of 264 pairs, which
constitutes roughly one-tenth of the original dataset of 2667 sense/antisense cluster
pairs, was selected for analysis by Microarrays (Table_S1 on CD-ROM2, an excerpt
of which is shown in Table 5 below). In this sample, the proportion of each of the
nine subgroups depicted in Table 4 is similar to that of the original dataset, indicating
a good representation of the various subgroups.

Table 4

mRNA/	No cluster	1 cluster	2 clusters	Total
Splicing	w introns	w intron(s)	w intron(s)	
No cluster w mRNA	48	132	197	377 (14%)
1 cluster w mRNA	17	490	1039	1546 (58%)
2 clusters w mRNA	1	85	658	744 (28%)
Total	66 (2.5%)	707 (26%)	1894 (71%)	2667 (100%)

123

Table represents the proportion of sense/antisense clusters in the dataset of 2667 that
contain: 1) a known mRNA and 2) expressed sequences spanning at least one intron, in
one of the two clusters, in both clusters or in none of the clusters.

Table 5 below is an excerpt of Table_S1 provided on CD-ROM2; Table 5
exemplifies five of the putative sense/antisense pairs that were selected for microarray
analysis. The first column provides the pair number. The next two columns provide
the accession numbers of representative expressed sequences from the overlapping
region of the sense and the antisense genes, respectively. The two columns identified
by the "RNA" header provide the accession numbers of known mRNAs in the sense
and antisense clusters (if available), and the last two columns provide the GenBank
descriptions of these mRNAs.

Table 5

Pair no.	sense seq. from over-lapping region	antisense seq. from overlapping region	RNA in sense cluster	RNA in a-sense cluster	description of RNA in sense cluster	description of RNA in antisense cluster
235	NM_ 6227	NM_ 308	NM_ 6227	NM_ 308	Homo sapiens phospholipid transfer protein (PLTP), mRNA #DV L26232.1	Homo sapiens protective protein for beta-galactosidase (galactosialidosis) (PPGB), mRNA
237	NM_ 4703	NM_ 2532	NM_ 4703	NM_ 2532	Homo sapiens rabaptin-5 (RAB5EP), mRNA #DV X91141.1	Homo sapiens nucleoporin 88kD (NUP88) mRNA #DV Y08612.2
217	NM_ 14885	AV 723808	NM_ 14885	NM_ 2940	Homo sapiens anaphase-promoting complex 10 (APC10) mRNA. #DV AL080090.1	Homo sapiens ATP-binding cassette, sub-family E (OABP), member 1 (ABCE1), mRNA.
209	BC 8865	BG 717574	NM_ 32231	NM_ 3099	Homo sapiens hypothetical protein FLJ22875 (FLJ22875), mRNA	Homo sapiens sorting nexin 1 (SNX1),mRNA. #DV U53225.1
196	BE 885605	AL 527611	NM_ 17832	NM_ 3640	Homo sapiens hypothetical protein FLJ20457 (FLJ20457), mRNA	Homo sapiens inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein (IKBKAP),mRNA

Table 5 Cont.

124

Microarrays were constructed by spotting each of the 264 pairs of oligonucleotide probes onto treated glass slides in quadruplicates. The two counterpart oligonucleotide probes of each pair were spotted next to each other to ensure similar hybridization conditions.

As positive controls, each of the blocks contained oligonucleotides spotted at various concentrations for four ubiquitously expressed housekeeping genes: guanine nucleotide binding protein beta polypeptide 2-like 1 (gnb2l1, HUMMHBA123, NM_006098), heat shock 70kD protein 10 (hsp70, HSHSC70CDS0, NM_006597), beta actin (actin, ACTB, NM_001101), and glyceraldehyde-3-phosphate dehydrogenase (gapdh, NM_002046).

Two random oligonucleotides were used as negative controls. These computer-generated arbitrary sequences displayed no alignment to human genome sequences but had the same physical characteristics as the other oligonucleotide probes. In addition, 22 probes for 11 previously documented sense/antisense pairs were also analyzed in the Microarrays (entries Pair no. "known 1"-"known 11" on Table_S1 of CD-ROM2).

The Microarrays were hybridized with poly(A)+ RNAs obtained from 19 human cell lines representing a variety of tissues and four normal human tissues (see *General Materials and Methods* section above). Each poly(A)+ RNA was reverse transcribed by priming with oligo(dT) and random nonamers, and engineered to incorporate a fluorescent marker. A pool containing an equal mix of the RNAs from all cell lines was also transcribed and used as a reference target. The resulting fluorescently-labeled cDNAs were combined and hybridized to the oligonucleotide Microarrays.

The experiments were performed in duplicate and utilized a fluorescent reversal of the Cy3- and Cy5-labelled cDNA. Stringent hybridization conditions were utilized in order to minimize the appearance of false positive signals, despite the possibility of compromised detection of low abundance transcripts.

The raw data was normalized at several levels; within each slide, between reciprocal slides, and globally between slides (see *General Materials and Methods* section above). Non-specific levels of hybridization were estimated from the negative

controls. The threshold for significant positive signals resulting from authentic hybridization was set at 4 standard deviations of the mean normalized signals for the negative controls. Processed data was presented as normalized signal intensity and as normalized signal ratios (Table_S2 on CD-ROM2).

To further substantiate array results, several pairs of oligonucleotides were also utilized in Northern blot analysis. Figures 22a-j illustrate results of such northern blot analysis. Figure 22a reveals expression patterns of randomly selected sequence pair number 235, denoted as Rand_235 in Table 6, below. Similarly, Figure 22b corresponds to pair number 173, Figure 22c to pair number 248, Figure 22d to pair number 6, Figure 22e to pair number 216, Figure 22f to pair number 239, Figure 22g to pair number 202, Figure 22h to pair number 114, Figure 22i to pair number 188, and Figure 22j to pair number 223. Eight pairs (Figures 22a-h) evaluated revealed positive signals for both sense and antisense expression, while two (Figures 22i-j) revealed a positive signal for only one of the genes, with the counterpart being a known RefSeq mRNA.

Figure 23 represents an excerpt of Table_S2 (provided in CD-ROM2) which summarizes the results obtained utilizing the array generated according to the teachings of the present invention. Expression thresholds were verified and indicated and normalization for microarray signals was conducted as described above. *Rji* ratios were obtained for each cell line/tissue assessed.

Taken cumulatively, the data presented herein revealed positive signals for both sense and antisense transcripts in 65 cluster pairs. In another 47 cases, significant hybridization signals were detected for antisense sequences with known counterpart sense transcripts, i.e. RefSeq mRNAs, which did not give clear hybridization signals on the Microarrays. Thus, 42.5 % (112 cases) of the 264 represented on the Microarrays, yielded detectable antisense transcription.

The conversion table, assigning the respective serial number as it appears in the "table_125" file of CD-ROM2 and "table_133" file of CD-ROM 3 enclosed herewith, is shown in Table 6 below.

Table 6

Rand_#	Serial No	Rand_#	Serial No	Rand_#	Serial No
Rand_1	2326	Rand_179	3266	Rand_258	3807

126

Rand_10	3647	Rand_18	3073	Rand_259	2621
Rand_100	2758	Rand_180	1794	Rand_26	4009
Rand_101	1595	Rand_181	1585	Rand_27	3393
Rand_102	3686	Rand_182	3554	Rand_28	3589
Rand_103	2331	Rand_183	3377	Rand_29	1837
Rand_104	3496	Rand_184	3466	Rand_3	3046
Rand_105	3134	Rand_185	3159	Rand_30	3297
Rand_106	1339	Rand_186	1413	Rand_31	3692
Rand_107	908	Rand_187	3645	Rand_32	707 2376
Rand_108	2929	Rand_188	3880	Rand_33	2052
Rand_109	2537	Rand_189	3009	Rand_34	1904
Rand_11	2806	Rand_19	3641	Rand_35	3718
Rand_110	3594	Rand_190	2549	Rand_36	3898
Rand_111	2819	Rand_191	2874	Rand_37	1821
Rand_112	3019	Rand_192	2515	Rand_38	3092
Rand_113	3815	Rand_193	3914	Rand_39	3262
Rand_114	2606	Rand_194	2751	Rand_4	3558
Rand_115	1662	Rand_195	2091	Rand_40	2474
Rand_116	2171	Rand_196	1966	Rand_41	3568
Rand_117	2539	Rand_197	3778	Rand_42	864
Rand_118	2802	Rand_198	3877	Rand_43	1864
Rand_119	2761	Rand_199	2248	Rand_44	3045
Rand_12	1947	Rand_2	3172	Rand_45	2854
Rand_120	3228	Rand_20	2360	Rand_46	3852
Rand_121	2076	Rand_200	2064	Rand_47	3096
Rand_122	1835	Rand_201	3597	Rand_48	1987
Rand_123	3029	Rand_202	2826	Rand_49	2893
Rand_124	2898	Rand_203	2388	Rand_5	2060
Rand_125	1568	Rand_204	3889	Rand_50	1058
Rand_126	2456	Rand_205	2211	Rand_51	3560
Rand_127	2019	Rand_206	3512	Rand_52	2604
Rand_128	2346	Rand_207	3452	Rand_53	3397
Rand_129	2460	Rand_208	3886	Rand_54	2040
Rand_13	2429	Rand_209	1600	Rand_55	3784
Rand_130	3374	Rand_21	2952	Rand_56	3659
Rand_131	3292	Rand_210	2432	Rand_57	2005 2688
Rand_132	3259	Rand_211	1651 3968	Rand_58	3187

Rand_133	3591	Rand_212	3074	Rand_59	1350
Rand_134	3340	Rand_213	2341	Rand_6	2202
Rand_135	1958	Rand_214	1984	Rand_60	3183
Rand_136	2274	Rand_215	2803	Rand_61	2275
Rand_137	3527	Rand_216	3806	Rand_62	3882
Rand_138	1533	Rand_217	2186	Rand_63	1044 3899
Rand_139	2622	Rand_218	857	Rand_64	2811
Rand_14	2058	Rand_219	1744	Rand_65	3232
Rand_140	2578	Rand_22	2285	Rand_66	3242
Rand_141	3492	Rand_220	2977	Rand_67	34 112 2727
Rand_142	3928	Rand_221	3863	Rand_68	3909
Rand_143	2282 3790	Rand_222	2846	Rand_69	4016
Rand_144	2820	Rand_223	3986	Rand_7	2337
Rand_145	1329	Rand_224	579 3688	Rand_70	2101 3707
Rand_146	1783	Rand_225	3984	Rand_71	3703
Rand_147	1527	Rand_226	2889	Rand_72	3477
Rand_148	2662	Rand_227	3869	Rand_73	2437
Rand_149	2031	Rand_228	3994	Rand_74	3808
Rand_15	2677	Rand_229	3818	Rand_75	3905
Rand_150	1303 1659	Rand_23	3890	Rand_76	1138 2194
Rand_151	1767	Rand_230	3152	Rand_77	819
Rand_152	3378	Rand_231	3445	Rand_78	3704
Rand_153	984	Rand_232	3663	Rand_79	2309
Rand_154	3759	Rand_233	3410	Rand_8	3441
Rand_155	2046	Rand_234	1112	Rand_80	1219
Rand_156	2528	Rand_235	3918	Rand_81	1416
Rand_157	283 1798 2048	Rand_236	2316	Rand_82	1543
Rand_158	3710	Rand_237	3673	Rand_83	3269
Rand_159	3178	Rand_238	3990	Rand_84	532 732
Rand_16	3336	Rand_239	4012	Rand_85	2607
Rand_160	1645	Rand_24	3250	Rand_86	1867
Rand_161	2074 3464	Rand_240	2932	Rand_87	627 3006
Rand_162	3436	Rand_241	3836	Rand_88	2068
Rand_163	2738	Rand_242	3424	Rand_89	2296
Rand_164	2749	Rand_243	3982	Rand_9	3741
Rand_165	2206	Rand_244	3472	Rand_90	1076
Rand_166	1349	Rand_245	2071	Rand_91	3385

128

Rand_167	2773	Rand_246	3904	Rand_92	2334
Rand_168	3305	Rand_247	2056	Rand_93	2833
Rand_169	1954	Rand_248	3855	Rand_94	2626
Rand_17	3940	Rand_249	2980	Rand_95	3671
Rand_170	2813	Rand_25	3453	Rand_96	1923
Rand_171	3868	Rand_250	3565	Rand_97	1863
Rand_172	762 1424 3942	Rand_251	2459	Rand_98	3437
Rand_173	3872	Rand_252	71 3147	Rand_99	3469
Rand_174	3801	Rand_253	3967	Rand_260	1975 3171
Rand_175	2547	Rand_254	702 2867 3088	Rand_261	4013
Rand_176	1251	Rand_255	3156	Rand_262	2418
Rand_177	1603	Rand_256	2324 2998	Rand_263	2451
Rand_178	2769	Rand_257	2284	Rand_264	3832

Table 6 Cont

Rand # = the name of the pair on the chip as it appears in Table_S2 on CD-ROM2, column "Probe"; Serial No = no of the pair in the Table files on CD-ROMs 2 and 3 (could be more than one in case the antisense event was separated to more than two contigs).

The sensitivity of the experimental approach utilized, i.e. the ability to detect a given transcript, stems from a combination of the stringency used in the microarray analysis and the level of expression and tissue specificity of the RNA. This can be estimated from the positive signals obtained for 65% of the oligos representing known RefSeq mRNAs on the Microarrays. This level of detection is comparable to that obtained in other studies, such as the 58% of known exons verified using microarray analysis (D. D. Shoemaker, et al., Nature 409, 922; 2001).

Thus, the present methodology provides a level of detection for a pair of genes that is 0.65 x 0.65= 0.42, a value supported by the detection of positive signals for both sense and antisense expression in 5 out of 11 (0.45) clusters of previously described sense/antisense pairs (Table_S2 on CD-ROM2).

Of the 264 cluster pairs analyzed in the Microarrays of the present invention, 65 clusters (0.25) showed significant signals for both sense and antisense transcripts, which is 60% of the proposed level of detection for a pair of genes (0.25/0.42). Extrapolating this figure to the predicted antisense dataset of 2667 clusters, predicts at least 1600 sense/antisense transcriptional units in the human genome.

129
EXAMPLE 10

Annotation of newly uncovered naturally occurring antisense transcripts

Newly uncovered naturally occurring transcripts were annotated using the Gencarta (Compugen, Tel-Aviv, Israel) platform. The Gencarta platform includes a rich pool of annotations, sequence information (particularly of spliced sequences), chromosomal information, alignments, and additional information such as SNPs, gene ontology terms, expression profiles, functional analyses, detailed domain structures, known and predicted proteins and detailed homology reports.

Brief description of the methodology used to obtain annotative sequence information is summarized infra (for detailed description see U.S. Pat. Appl. 10/426,002).

The ontological annotation approach - An ontology refers to the body of knowledge in a specific knowledge domain or discipline such as molecular biology, microbiology, immunology, virology, plant sciences, pharmaceutical chemistry, medicine, neurology, endocrinology, genetics, ecology, genomics, proteomics, cheminformatics, pharmacogenomics, bioinformatics, computer sciences, statistics, mathematics, chemistry, physics and artificial intelligence.

An ontology includes domain-specific concepts – referred to, herein, as sub-ontologies. A sub-ontology may be classified into smaller and narrower categories. The ontological annotation approach is effected as follows.

First, biomolecular (i.e., polynucleotide or polypeptide) sequences are computationally clustered according to a progressive homology range, thereby generating a plurality of clusters each being of a predetermined homology of the homology range.

Progressive homology is used to identify meaningful homologies among biomolecular sequences and to thereby assign new ontological annotations to sequences, which share requisite levels of homologies. Essentially, a biomolecular sequence is assigned to a specific cluster if displays a predetermined homology to at least one member of the cluster (i.e., single linkage). A "progressive homology range" refers to a range of homology thresholds, which progress via predetermined increments from a low homology level (e.g. 35 %) to a high homology level (e.g. 99 %).

130

Following generation of clusters, one or more ontologies are assigned to each cluster. Ontologies are derived from an annotation preassociated with at least one biomolecular sequence of each cluster; and/or generated by analyzing (e.g., text-mining) at least one biomolecular sequence of each cluster thereby annotating biomolecular sequences.

The hierarchical annotation approach - "Hierarchical annotation" refers to any ontology and subontology, which can be hierarchically ordered, such as, a tissue expression hierarchy, a developmental expression hierarchy, a pathological expression hierarchy, a cellular expression hierarchy, an intracellular expression hierarchy, a taxonomical hierarchy, a functional hierarchy and so forth.

The hierarchical annotation approach is effected as follows.

First, a dendrogram representing the hierarchy of interest is computationally constructed. A "dendrogram" refers to a branching diagram containing multiple nodes and representing a hierarchy of categories based on degree of similarity or number of shared characteristics.

Each of the multiple nodes of the dendrogram is annotated by at least one keyword describing the node, and enabling literature and database text mining, such as by using publicly available text mining software. A list of keywords can be obtained from the GO Consortium (www.geneontlogy.org). However, measures are taken to include as many keywords, and to include keywords which might be out of date. For example, for tissue annotation, a hierarchy is built using all available tissue/libraries sources available in the GenBank, while considering the following parameters: ignoring GenBank synonyms, building anatomical hierarchies, enabling flexible distinction between tissue types (normal versus pathology) and tissue classification levels (organs, systems, cell types, etc.).

In a second step, each of the biomolecular sequences is assigned to at least one specific node of the dendrogram.

The biomolecular sequences can be annotated biomolecular sequences, unannotated biomolecular sequences or partially annotated biomolecular sequences.

Annotated biomolecular sequences can be retrieved from pre-existing annotated databases as described hereinabove.

131

For example, in GenBank, relevant annotational information is provided in the definition and keyword fields. In this case, classification of the annotated biomolecular sequences to the dendrogram nodes is directly effected. A search for suitable annotated biomolecular sequences is performed using a set of keywords which are designed to classify the biomolecular sequences to the hierarchy (i.e., same keywords that populate the dendrogram)

In cases where the biomolecular sequences are unannotated or partially annotated, extraction of additional annotational information is effected prior to classification to dendrogram nodes. This can be effected by sequence alignment, as described hereinabove. Alternatively, annotational information can be predicted from structural studies. Where needed, nucleic acid sequences can be transformed to amino acid sequences to thereby enable more accurate annotational prediction.

Finally, each of the assigned biomolecular sequences is recursively classified to nodes hierarchically higher than the specific nodes, such that the root node of the dendrogram encompasses the full biomolecular sequence set, which can be classified according to a certain hierarchy, while the offspring of any node represent a partitioning of the parent set.

For example, a biomolecular sequence found to be specifically expressed in "rhabdomyosarcoma", will be classified also to a higher hierarchy level, which is "sarcoma", and then to "Mesenchimal cell tumors" and finally to a highest hierarchy level "Tumor". In another example, a sequence found to be differentially expressed in endometrium cells, will be classified also to a higher hierarchy level, which is "uterus", and then to "women genital system" and to "genital system" and finally to a highest hierarchy level "genitourinary system". The retrieval can be performed according to each one of the requested levels.

Annotating gene expression according to relative abundance - Spatial and temporal gene annotations are also assigned by comparing relative abundance in libraries of different origins. This approach can be used to find gene which are differentially expressed in tissues, pathologies and different developmental stages. In principal, the presentation of a contig in at least two tissues of interest is determined and significant over or under representation of the contig in one of the at least two

tissues is assessed to identify differential expression. Significant over or under representation is analyzed by statistical pairing.

Annotating spatial and temporal expression can also be effected on splice variants. This is effected as follows. First, a contigue which includes exonal sequence presentation of the at least two splice variants of the gene of interest is obtained. This contigue is assembled from a plurality of expressed sequences;

Then, at least one contigue sequence region unique to a portion (i.e., at least one and not all) of the at least two splice variants of the gene of interestis identified . Identification of such unique sequence region is effected using computer alignment software.

Finally, the number of the plurality of expressed sequences in the tissue having the at least one contigue sequence region is compared with the number of the plurality of expressed sequences not-having the at least one contigue sequence region, to thereby compare the expression level of the at least two splice variants of the gene of interest in the tissue.

Sequence anntotations obtained using the above-described methodologies and other approaches are disclosed in a data table in the pep_seqs_133 of the enclosed CD-ROM3.

The data table shows a collection of annotations for biomolecular sequences, which were identified according to the teachings of the present invention using transcript data based on GenBank versions 136.

Each feature in the data table is identified by "#".

#INDICATION – This field designates the indications (i.e., diseases, disorders, pathological conditions) and therapies that the polypeptide of the present invention can be utilized for. Specifically, an indication lists the disorders or diseases in which the polypeptide of the present invention can be clinically used. A therapy describes a postulated mode of action of the polypeptide for the above-mentioned indication. For example, an indication can be "Cancer, general" while the therapy will be "Anticancer".

Each protein was assigned a SWISSPROT and/or TremBl human protein accession as described in section "Assignment of Swissprot/TremBl accessions to Gencarta contigs" hereinbelow. The information contained in this field is the

133

indication concatenated to the therapies that were accumulated for the SWISSPROT and/or TremBl human protein from drug databases, such as PharmaProject (PJB Publications Ltd 2003 www.pjbpubs.com/cms.asp?pageid=340) and public databases, such as LocusLink (www.genelynx.org/cgi-bin/resource?res=locuslink) and Swissprot (http://www.ebi.ac.uk/swissprot/index.html). The field may comprise more than one term wherein a ";" separates each adjacent terms.

Example- #INDICATION Alopecia, general; Antianginal; Anticancer, immunological; Anticancer, other; Atherosclerosis; Buerger's syndrome; Cancer, general; Cancer, head and neck; Cancer, renal; Cardiovascular; Cirrhosis, hepatic; Cognition enhancer; Dermatological; Fibrosis, pulmonary; Gene therapy; Hepatic dysfunction, general; Hepatoprotective; Hypolipaemic/Antiatherosclerosis; Infarction, cerebral; Neuroprotective; Ophthalmological; Peripheral vascular disease; Radio/chemoprotective; Recombinant growth factor; Respiratory; Retinopathy, diabetic; Symptomatic antidiabetic; Urological;

Assignment of Swissprot/TremBl accessions to Gencarta contigs - Gencarta contigs were assigned a Swissprot/TremBl human accession as follows. Swissprot/TremBl data were parsed and for each Swissprot/TremBl accession (excluding Swissprot/TremBl that are annotated as partial or fragment proteins) cross-references to EMBL and Genbank were parsed. The alignment quality of the Swissprot/TremBl protein to their assigned mRNA sequences was checked by frame+p2n alignment analysis. A good alignment was considered as heving the following properties:

• For partial mRNAs (those that in the mRNA description have the phrase "partial cds" or annotated as "3'" or "5'")- an overall identity of 97% and coverage of 80 % of the Swissprot/TremBl protein.

• All the rest were considered as full coding mRNAs and for them an overall identity of 97% identity and coverage of the Swissprot/TremBl protein of over 95 %.

The mRNAs were searched in the LEADS database for their corresponding contigs, and the contigs that included these mRNA sequences were assigned the Swissprot/TremBl accession.

#PHARM- This field indicates possible pharmacological activities of the polypeptide. Each polypeptide was assigned with a SWISSPROT and/or TremBl human protein accession, as described above. The information contained in this field is the indication concatenated to the therapies that were accumulated for the SWISSPROT and/or TremBl human protein from drug databases such as PharmaProject (PJB Publications Ltd 2003 www.pjbpubs.com/cms.asp?pageid=340) and public databases, such as LocusLink and Swissprot. Note that in some cases this field can include opposite terms in cases where the protein can have contradicting activities – such as:

(i)   Stimulant – inhibitor;

(ii)   Agonist – antagonist;

(iii)   Activator- inhibitor;

(iv)   Immunosuppressant – Immunostimulant;

In these cases the pharmacology was indicated as "modulator". For example, if the predicted polypeptide has potential agonistics/antagonistic effects (e.g. Fibroblast growth factor agonist and Fibroblast growth factor antagonist) then the annotation for this code will be "Fibroblast growth factor modulator"

A documentated example for such contradicing activities has been described for the soluble tumor necrosis factor receptors [Mohler et al., J. Immunology 151, 1548-1561]. Essentially, Mohler and co-workers showed that soluble receptor can act both as a carrier of TNF (i.e., agonistic effect) and as an antagonist of TNF activity.

#THERAPEUTIC_PROTEIN – This field predicts a therapeutic role for a protein represented by the contig. A contig was assigned this field if there was information in the drug database or the public databases (e.g., described hereinabove) that this protein, or part thereof, is used or can be used as a drug. This field is accompanied by the swissprot accession of the therapeutic protein which this contig most likely represents. Example: # THERAPEUTIC_PROTEIN UROK_HUMAN

GO annotations were predicted as described in "The ontological annotation approach" section hereinabove. Functional annotations of transcripts based on Gene Ontology (GO) are indicated by the following format.

*, ** "#GO_P", annotations related to Biological Process,

135

*, ** "#GO_F", annotations related to Molecular Function, and

*, ** "#GO_C", annotations related to Cellular Component.

Proloc was used for protein subcellular localization prediction that assigns GO cellular component annotation to the protein. The localization terms were assigned with GO entries.

Cellular localization – ProLoc software, commercially available from Compugen LTD, was used to predict the cellular localization of the proteins. Two main approaches were used: (i) the presence of known extracellular domain/s in a protein; (ii) calculating putative transmembrane segments, if any, in the protein and calculating 2 p-values for the existence of a signal peptide. The latter is done by a searching for a signal peptide at the N-terminal sequence of the protein generating a score. Running the program on real signal peptides and on N-terminal protein sequences that lack a signal peptide resulted in 2 score distributions: the first is the score distribution of the real signal peptides and the second is the score distribution of the N-terminal protein sequences that lack the signal peptide. Given a novel protein product, ProLoc calculates the above-score score and provides the percentage of the scores that are higher than the current score, in the first distribution, as a first p-value (lower p-values mean more reliable signal peptide prediction) and the percentage of the scores that are lower than the current score, in the second distribution, as a second p-value (lower p-values mean more reliable non signal peptide prediction).

Thus, using this algorithm secreted proteins and membrane proteins can be identified, for example. However, proteins which lack signal peptide while are still secreted (such as after lysis of viral infected cells) can be identified such as by homology search to extracellular proteins which were identified as such by ProLoc.

Table 7

IPR000874	Bombesin-like peptide
IPR001693	Calcitonin-like
IPR001651	Gastrin/cholecystokinin peptide hormone
IPR000532	Glucagon/GIP/secretin/VIP
IPR001545	Gonadotropin, beta chain
IPR004825	Insulin/IGF/relaxin
IPR000663	Natriuretic peptide
IPR001955	Pancreatic hormone
IPR001400	Somatotropin hormone
IPR002040	Tachykinin/Neurokinin

IPR006081	Alpha defensin
IPR001928	Endothelin-like toxin
IPR001415	Parathyroid hormone
IPR001400	Somatotropin hormone
IPR001990	Chromogranin/secretogranin
IPR001819	Chromogranin A/B
IPR002012	Gonadotropin-releasing hormone
IPR001152	Thymosin beta-4
IPR000187	Corticotropin-releasing factor, CRF
IPR001545	Gonadotropin, beta chain
IPR000476	Glycoprotein hormones alpha chain
IPR000476	Glycoprotein hormones alpha chain
IPR001323	Erythropoietin/thrombopoeitin
IPR001894	Cathelicidin
IPR001894	Cathelicidin
IPR001483	Urotensin II
IPR006024	Opioid neuropeptide precursor
IPR000020	Anaphylatoxin/fibulin
IPR000074	Apolipoprotein A1/A4/E
IPR001073	Complement C1q protein
IPR000117	Kappa casein
IPR001588	Casein, alpha/beta
IPR001855	Beta defensin
IPR001651	Gastrin/cholecystokinin peptide hormone
IPR000867	Insulin-like growth factor-binding protein, IGFBP
IPR001811	Small chemokine, interleukin-8 like
IPR004825	Insulin/IGF/relaxin
IPR002350	Serine protease inhibitor, Kazal type
IPR000001	Kringle
IPR002072	Nerve growth factor
IPR001839	Transforming growth factor beta (TGFb)
IPR001111	Transforming growth factor beta (TGFb), N-terminal
IPR001820	Tissue inhibitor of metalloproteinase
IPR000264	Serum albumin family
IPR005817	Wnt superfamily

For each category the following features are optionally addressed:

"#GO_Acc" represents the accession number of the assigned GO entry, corresponding to the following "#GO_Desc" field.

"#GO_Desc" represents the description of the assigned GO entry, corresponding to the mentioned "#GO_Acc" field.

"#CL" represents the confidence level of the GO assignment, when #CL1 is the highest and #CL5 is the lowest possible confidence level. This field appears only when the GO assignment is based on a Swissprot/TremBl protein accession or

137

Interpro accession and (not on Proloc predictions or viral proteins predictions). Preliminary confidence levels were calculated for all public proteins as follows:

PCL 1: a public protein that has a curated GO annotation,

PCL 2: a public protein that has over 85 % identity to a public protein with a curated GO annotation,

PCL 3: a public protein that exhibits 50 - 85 % identity to a public protein with a curated GO annotation,

PCL 4: a public protein that has under 50 % identity to a public protein with a curated GO annotation.

For each protein a homology search against all public proteins was done. If the protein has over 95 % identity to a public protein with PCL X then the protein gets the same confidence level as the public protein. This confidence level is marked as "#CL X". If the protein has over 85 % identity but not over 95 % to a public protein with PCL X than the protein gets a confidence level lower by 1 than the confidence level of the public protein. If the protein has over 70 % identity but not over 85 % to a public protein with PCL X than the protein gets a confidence level lower by 2 than the confidence level of the public protein. If the protein has over 50 % identity but not over 70 % to a public protein with PCL X than the protein gets a confidence level lower by 3 than the confidence level of the public protein. If the protein has over 30 % identity but not over 50 % to a public protein with PCL X than the protein gets a confidence level lower by 4 than the confidence level of the public protein.

A protein may get confidence level of 2 also if it has a true interpro domain that is linked to a GO annotation www.geneontology.org/external2go/interpro2go/.

When the confidence level is above "1", GO annotations of higher levels of the GO hierarchy are assigned (e.g. for "#CL 3" the GO annotations provided, is as appears plus the 2 GO annotations above it in the hierarchy).

"#DB" marks the database on which the GO assignment relies on. The "sp", relates to SwissProt/TremBl Protein knowledgebase, available from www.expasy.ch/sprot/. "InterPro", refers to the InterPro combined database, available from www.ebi.ac.uk/interpro/, which contains information regarding protein families, collected from the following databases: SwissProt (http://www.ebi.ac.uk/swissprot/), Prosite (http://www.expasy.ch/prosite/), Pfam

138

(www.sanger.ac.uk/Software/Pfam/),   Prints
(www.bioinf.man.ac.uk/dbbrowser/PRINTS/),   Prodom
(prodes.toulouse.inra.fr/prodom/), Smart (http://smart.embl-heidelberg.de/) and
Tigrfams (http://www.tigr.org/TIGRFAMs/). "Proloc statistical database"- meaning
the statistics Proloc uses for predicting the subcellulat localization of a protein.

"#EN" represents the accession of the entity in the database (#DB),
corresponding to the accession of the protein/domain why the GO was predicted. If
the GO assignment is based on a protein from the SwissProt/TremBl Protein database
this field will have the locus name of the protein. Examples, *"#DB sp #EN
NRG2_HUMAN"* means that the GO assignment in this case was based on a protein
from the SwissProt/Trembl database, while the closest homologue (that has a GO
assignment) to the assigned protein is depicted in SwissProt entry "NRG2_HUMAN
"#DB interpro #EN IPR001609" means that GO assignment in this case was based on
InterPro database, and the protein had an Interpro domain, *IPR001609,* that the
assigned GO was based on. In Proloc predictions this field will have a Proloc
annotation "#EN Proloc". In predicitions based on viral proteins this field will have
the gi. viral protein accession, "#EN 1491997".

#GENE_SYMBOL – for each Gencarta contig a HUGO gene symbol was
assigned in two ways:

(i)   After assigning a Swissprot/TremBl protein to each contig (see
Assignment of Swissprot/TremBl accessions to Gencarta contigs) all the gene symbols
that appear for the Swissprot entry were parsed and added as a Gene symbol
annotation to the gene.

(ii)   LocusLink information- LocusLink was downloaded from NCBI
ftp.ncbi.nih.gov/refseq/LocusLink/ (files loc2acc, loc2ref, and LL.out_hs). The data
was integrated producing a file containing the gene symbol for every sequence.
Gencarta contigs were assigned a gene symbol if they contain a sequence from this
file that has a gene symbol

Example: #GENE_SYMBOL MMP15

#DIAGNOSTICS- secreted/membranal proteins get an annotation of "can be
used as diagnostic markers for" for the list of indications as appearing in the #
INDICATION field, described hereinabove. All proteins that were identified as

secreted or membranal proteins (as described in the GO field section) will be assigned with this field.

In addition, known Gencarta contigs representing known diagnostic markers (such as listed in Table 8, below) and all transcripts and proteins deriving from this contig will be assigned to this field and will get the above mentioned annotation followed by "as indicated in the Diagnostic markers table".

Table 8

Enzymes		
Test	*Gencarta Contig*	*Comments*
GPT	R35137 (GPT glutamic-pyruvate transaminase (alanine aminotransferase)) Z24841 (GPT2 glutamic pyruvate transaminase (alanine aminotransferase) 2)	Also called ALT – alanine aminotransferase. Standard liver function test
GOT	M78228 (GOT1 glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1)) M86145 (GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	Also called AST – aspartate aminotransferase. Standard liver function test
GGT	<u>HUMGGTX</u> (GGT1: gamma-glutamyltransferase 1)	Liver disease
CPK	T05088 (CKB creatine kinase, brain) HUMCKMA (CKM creatine kinase, muscle) H20196 (CKMT1 creatine kinase, mitochondrial 1 (ubiquitous)) HUMSMCK (CKMT2 creatine kinase, mitochondrial 2 (sarcomeric))	Also called CK. Mostly used for muscle pathologies. The MB variant is heart specific and used in the diagnosis of myocardial infarction
CPK-MB	T05088 (CKB creatine kinase, brain) HUMCKMA (CKM creatine kinase, muscle)	Cardiac problems – hetro-dimer of CKB and CKM
Alkaline Phosphatase	HSAPHOL- ALPL: alkaline phosphatase, liver/bone/kidney HUMALPHB - ALPI: alkaline phosphatase, intestinal HUMALPP- ALPP: alkaline phosphatase, placental (Regan isozyme)	Bone related syndromes and liver diseases, mostly with biliary involvement
Amylase	AA367524- (AMY1A: amylase, alpha 1A; salivary) T10898- (AMY2B: amylase, alpha 2B; pancreatic and 2A)	Blood/Urine. Pancreas related diseases
LDH	HSLDHAR (LDHA lactate dehydrogenase A) M77886 (LDHB lactate dehydrogenase B) HSU13680 (LDHC lactate dehydrogenase C) AA398148 (LDHL lactate dehydrogenase A –like)	Lactate Dehydrogenase. Used for myocardial infarction diagnosis and neoplastic syndromes assessment.

	R09053 (LDHD lactate dehydrogenase D)	
G6PD	S58359 (G6PD glucose-6-phosphate dehydrogenase)	Glucose 6-phosphate dehydrogenase. Levels measured when deficiency is suspected (leading to susceptibility to hemolysis)
Alpha1 antiTrypsin	HUMA1ACM (SERPINA3 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3) T10891 (AGT angiotensinogen (serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), memb er 8)) R83168 (SERPINA6 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6) HUMCINHP (SERPINA5 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5) HSA1ATCA (SERPINA1 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1) HUMKALLS (SERPINA4 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 4) HUMTBG (SERPINA7 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7) T60354 (SERPINA10 serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10)	Chronic lung diseases
Renin	HSRENK (REN renin)	Some hypertension syndromes
Acid Phosphatase	HUMAAPA (ACP1: acid phosphatase 1, soluble) T48863 (ACP2: acid phosphatase 2, lysosomal) HSMRACP5 (ACP5: acid phosphatase 5, tartrate resistant) T85211 (ACP6: lysophosphatidic acid phosphatase) HSPROSAP (ACPP: acid phosphatase, prostate) AA005037 (ACPT: acid phosphatase, testicular)	Used to differentiate multiple myeloma with other monoclonal gammopathies of uncertain significance
Beta glucoronidase	T11069 (GUSB glucuronidase, beta)	Used to differentiate multiple myeloma with other monoclonal gammopathies of uncertain significance
Aldolase	HSALDAR (ALDOA aldolase A, fructose-bisphosphate)	Glycogen storage diseases

141

	HSALDOBR (ALDOB aldolase B, fructose-bisphosphate) M62176 (ALDOC aldolase C, fructose-bisphosphate)	
Choline esterase	HUMCHEF (BCHE butyrylcholinesterase) F00931 (ACHE acetylcholinesterase (YT blood group))	Probably used for organophosphates/"nerve gases" intoxications
Pepsinogen	HUMPGCA PGC: progastricsin (pepsinogen C)	(in the stomach), high in gastritis, low in pernicious anemia[
ACE	HSACE (ACE: angiotensin I converting enzyme (peptidyl-dipeptidase A) 1) AA397955 (ACE2: angiotensin I converting enzyme (peptidyl-dipeptidase A) 2)	Angiotensin-converting enzyme. Sarcoidosis

Miscelleneous		
Test	*Gencarta Contig*	*Comments*
Prion Protein	HUMPRP0A (PRNP prion protein (p27-30) (Creutzfeld-Jakob disease, Gerstmann-Straus ler-Scheinker syndrome, fatal familial insomnia)) W73057 (PRND prion protein 2 (dublet))	BSE diagnosis
Myelin basic protein	M78010 (MBP myelin basic protein) R13982 (MOBP myelin-associated oligodendrocyte basic protein)	In CSF. In Multiple sclerosis
Albumin	HSALB1 (ALB albumin)	Mostly liver function and failure of intestine absorption
Prealbumin	HSALB1 (ALB albumin)	early diagnosis of malabsorption
Ferritin	HUMFERLS (FTL ferritin, light polypeptide) HUMFERHA (FTH1 ferritin, heavy polypeptide 1)	Iron deficiency anemia
Transferrin	S95936 (TF transferrin)	Iron deficiency anemia
Haptoglobin	HUMHPA1B (HP haptoglobin)	Used in anemia states and neoplastic syndromes
CRP	HSCREACT (CRP C-reactive protein, pentraxin-related)	C reactive protein. Associated with active inflammation
AFP	D11581 (AFP alpha-fetoprotein)	Alpha Feto Protein. Used in pregnancy for abnormalities screening and as a cancer marker.
C3	T40158 (C3 complement component 3)	Various auto-immune and allergy syndromes
C4	HSCOC4 (C4A complement component 4A; C4B complement component 4B)	Various auto-immune and allergy syndromes
Ceruloplasmin	HSCP2 (CP ceruloplasmin (ferroxidase))	Wilson's disease (liver disease)
Myoglobin	T11628 (MB myoglobin)	Rhabdomyolysis, Myocardial infarction
FABP	S67314 (FABP3: fatty acid binding protein 3, muscle and heart)	myoglobin and Fatty Acid Binding

	D11754 (**FABP1** liver- L-FABP- fatty acid binding protein 1) AW605378 (FABP2: fatty acid binding protein 2, intestinal) HUMALBP (FABP4: fatty acid binding protein 4, adipocyte) T06152 (FABP5: fatty acid binding protein 5 (psoriasis-associated) HSI15PGN1 (FABP6: fatty acid binding protein 6, ileal (gastrotropin) R60348 (FABP7: fatty acid binding protein 7, brain)	
Troponin I	HUMTROPNIN (TNNI2 troponin I, skeletal, fast) Z25083 (TNNI1 troponin I, skeletal, slow) HUMTROPIA (TNNI3 troponin I, cardiac)	Acute myocardial infarction
Beta-2-microglobulin	HSB2MMU (B2M beta-2-microglobulin)	
Macroglobin	M62177 (A2M: alpha-2-macroglobulin)	Elevated in inflammation
Alpha-1 glycoprotein	T72188 (A1BG: alpha-1-B glycoprotein)	Elevated in inflammation and tumors.
Apo A-I	HUMAPOAIP (APOA1: apolipoprotein A-I)	Risk for coronary artery disease
Apo B-100	HSAPOBR2 (APOB: apolipoprotein B (including Ag(x) antigen))	Atherosclerotic heart disease
Apo E	T61627 (APOE: apolipoprotein E)	diagnosis of Type III hyperlipoproteinemia, evaluate a possible genetic component to atherosclerosis, or to help confirm a diagnosis of late onset AD
CF gene	HUMCFTRM (CFTR: cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7))	Cystic fibrosis disease (a DNA test – blood sample)
PSEN1 gene	T89701 (PSEN1: presenilin 1 (Alzheimer disease 3))	Early onset of familial AD (a DNA test – blood sample)
Hormones		
Test	***Gencarta Contig***	***Comments***
Erythropoietin	HSERPR (EPO erythropoietin)	Hardly used for diagnosis. Used as treatment
GH	HSGROW1 (GH1 growth hormone 1) HUMCS2 (GH2 growth hormone 2)	Growth Hormone. Endocrine syndromes
TSH	AV745295 (TSHB thyroid stimulating hormone, beta)	Part of thyroid functions tests
BetaHCG	R27266 (CGB5 chorionic gonadotropin, beta polypeptide 5)	Pregnancy, malignant syndromes in men and women
LH	HUMCGBB50 (LHB luteinizing hormone beta polypeptide)	Part of standard hormonal profile for fertility, gynecological syndromes and endocrine syndromes
FSH	AV754057 (FSHB follicle stimulating hormone, beta	Part of standard hormonal profile for fertility, gynecological syndromes and

	polypeptide)	endocrine syndromes
TBG	S40807 (TG thyroglobulin)	Thyroxin binding globulin. Thyroid syndromes
Prolactin	HSLACT (PRL prolactin)	Various endocrine syndromes
Thyroglobulin	S40807 (TG thyroglobulin)	Follow up of thyroid cancer patients
PTH	HSTHYR (PTH parathyroid hormone)	Parathyroid Hormone. Syndromes of calcium management
Insulin/Pre Insulin	HSPPI (INS insulin)	Diabetes
Gastrin	HSGAST (GAS gastrin)	Peptic ulcers
Oxytocin	HUMOTCB (OXT oxytocin, prepro-(neurophysin I))	Endocrine syndromes related to lactation
AVP	HUMVPC (AVP arginine vasopressin (neurophysin II, antidiuretic hormone, diabetes insipidus, neurohypophyseal))	Arginine Vasopressin. Endocrine syndromes related to the osmotic pressure of body fluids
ACTH	HUMPOMCMTC (POMC: proopiomelanocortin (adrenocorticotropin/ beta-lipotropin/ alpha-melanocyte stimulating hormone/ beta-melanocyte stimulating hormone/ beta-endorphin))	Secreted from the anterior pituitary gland. Regulation of cortisol. Abnormalities are indicative of Cushing's disease, addison's disease and adrenal tumors
BNP	HUMNATPEP (NPPB: natriuretic peptide precursor B)	Heart failure

Blood Clotting		
Test	*Gencarta Contig*	*Comments*
Protein C	S50739 (PROC protein C (inactivator of coagulation factors Va and VIIIa))	Inherited Clotting disorders
Protein S	HSSPROTR (PROS1 protein S (alpha))	Inherited Clotting disorders
Fibrinogen	D11940 (FGA: fibrinogen, A alpha polypeptide) HUMFBRB (FGB: fibrinogen, B beta polypeptide) T24021 (FGG: fibrinogen, gamma polypeptide)	Clotting disorders
Factors 2, 5, 7, 9, 10, 11, 12, 13	HUMPTHROM (F2 coagulation factor II (thrombin)) HUMTFPC (F3 coagulation factor III (thromboplastin, tissue factor)) HUMF5A (F5 coagulation factor V (proaccelerin, labile factor)) M78203 (F7 coagulation factor VII (serum prothrombin conversion accelerator)) HUMF8C (F8 coagulation factor VIII, procoagulant component (hemophilia A)) HUMCFIX (F9 coagulation factor IX (plasma thromboplastic component, Christmas dis ease, hemophilia B)) HUMCFX (F10: coagulation factor X) HUMFXI (F11 coagulation factor XI (plasma thromboplastin antecedent)) HUMCFXIIA (F12 coagulation factor XII (Hageman factor)) HUMFXIIIA (F13A1 coagulation factor XIII,	Inherited Clotting disorders

144

	A1 polypeptide) R28976 (F13B coagulation factor XIII, B polypeptide)	
vWF	HUMVWF (VWF von Willebrand factor)	Von Willebrand factor. Inherited Clotting disorders
Antithrombin III	T62060 (SERPINC1 serine (or cysteine) proteinase inhibitor, clade C (antithrombin), member 1)	Inherited Clotting disorders

Cancer Markers

Test	Gencarta Contig	Comments
AFP	D11581 (AFP alpha-fetoprotein)	Pregnancy, testicular cancer and hepatocellular cancer
CA125	HSIAI3B (M17S2 membrane component, chromosome 17, surface marker 2 (ovarian carcinoma antigen CA125))	Ovarian cancer
CA-15-3	HSMUC1A (MUC1 mucin 1, transmembrane)	Breast cancer
CA-19-9	HSAFUTF (FUT3: fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group included))	Gastrointestinal cancer, pancreatic cancer
CEA	T10888 HUMCEA (CEACAM3 carcinoembryonic antigen-related cell adhesion molecule 3)	Carcinoembryonic Antigen. Colorectal cancer
PSA	HSCDN9 (KLK3: kallikrein 3, (prostate specific antigen))	
PSMA	HUMPSM (FOLH1: folate hydrolase (prostate-specific membrane antigen) 1)	
TPA, TATI, OVX1, LASA, CA54/81	HSPSTI (SPINK1: serine protease inhibitor, Kazal type 1)	Ovarian cancer
BRCA 1	H90415 (BRCA1: breast cancer 1, early onset)	
BRCA 2	H47777 (BRCA2: breast cancer 2, early onset)	Breast cancer (ovarian cancer?)
HER2/Neu	S57296 (ERBB2: v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian))	Breast cancer
Estrogen receptor	HSERG5UTA (ESR1: estrogen receptor 1) HSRNAERB (ESR2: estrogen receptor 2 (ER beta))	Breast cancer
Progesterone receptor	T09102 (PGRMC1: progesterone receptor membrane component 1) Z32891 (PGRMC2: progesterone receptor membrane component 2)	Breast cancer

Note:

    (i)   Small portion of these "markers" are also drug targets, whether already for approved drugs (such as alpha1 antiTrypsin) or under development (e.g., GOT).

    (ii)   Some of these "markers" are also used as therapeutic proteins (e.g., Erythropoietin).

5   (iii)   All markers are found in the blood/serum unless otherwise specified.

#DRUG_DRUG_INTERACTION: refers to proteins involved in a biological process which mediates the interaction between at least two consumed drugs. Novel splice variants of known proteins involved in interaction between drugs may be used, for example, to modulate such drug-drug interactions. Examples of proteins involved in drug-drug interactions are presented in Table 9 together with the corresponding internal gene contig name, enabling to allocate the new splice variants within the data files in the attached CD-ROMs 1 and 3-PLEASE COMPLETE.

Table 9

Contig	Gene Symbol	Description
HUMANTLA	SLC3A2	4f2 cell-surface antigen heavy chain
Z43093	HTR6	5-hydroxytryptamine 6 receptor
HSXLALDA	ABCD1	Adrenoleukodystrophy protein
R35137	GPT	Alanine aminotransferase
D11683	ALDH1	Aldehyde dehydrogenase, cytosolic
T53833	AOX1	Aldehyde oxidase
HUMD4G08M3	ORM1	Alpha-1-acid glycoprotein 1
HUMD4G08M3	ORM2	Alpha-1-acid glycoprotein 2
HUMABPA	ABP1	Amiloride-sensitive amine oxidase [copper-containing]
S62734	MAOB	Amine oxidase [flavin-containing] b
AA526963	SLC6A14	Amino acid transporter b0+
HSAE2	SLC4A2	Anion exchange protein 2
M78110	SLC4A3	Anion exchange protein 3
M78052	ABCB2	Antigen peptide transporter 1
HUMMHCIIAB	ABCB3	Antigen peptide transporter 2
F02693	APOD	Apolipoprotein d
M62234	ASNA1	Arsenical pump-driving ATPase
HUMNORTR	NAT1	Arylamine n-acetyltransferase 1
T67129	NAT1	Arylamine n-acetyltransferase 1
AI262683	NAT2	Arylamine n-acetyltransferase 2
Z39550	ABCB9	ATP-binding cassette protein abcb9
Z44377	ABCA1	ATP-binding cassette, sub-family a, member 1
M78056	ABCA2	ATP-binding cassette, sub-family a, member 2
T05334	ABCA3	ATP-binding cassette, sub-family a, member 3
T79973	ABCB6	ATP-binding cassette, sub-family b, member 6, mitochondrial
T78010	ABCB7	ATP-binding cassette, sub-family b, member 7, mitochondrial
R89046	ABCB8	ATP-binding cassette, sub-family b, member 8, mitochondrial
H64439	ABCD2	ATP-binding cassette, sub-family d, member 2
M85760	ABCD3	ATP-binding cassette, sub-family d, member 3
Z21904	ABCD4	ATP-binding cassette, sub-family d, member 4
Z39977	ABCG1	ATP-binding cassette, sub-family g, member 1
Z45628	ABCG2	ATP-binding cassette, sub-family g, member 2
T80665	SLC7A9	B(0,+)-type amino acid transporter 1
AF091582	ABCB11	Bile salt export pump
Z38696	BLMH	Bleomycin hydrolase
T08127	BNPI	Brain-specific na-dependent inorganic phosphate cotransporter
F00545	SLC12A2	Bumetanide-sensitive sodium-(potassium)-chloride cotransporter 2
HSU07969	CDH17	Cadherin-17
T10238	SLC25A12	Calcium-binding mitochondrial carrier protein aralar1
Z40674	SLC25A13	Calcium-binding mitochondrial carrier protein aralar2
T61818	ABCC2	Canalicular multispecific organic anion transporter 1

T39953	ABCC3	Canalicular multispecific organic anion transporter 2
HUMCRE	CBR1	Carbonyl reductase [nadph] 1
AA320697	CBR3	Carbonyl reductase [nadph] 3
F03362	COMT	Catechol o-methyltransferase, membrane-bound form
T11004	COMT	Catechol o-methyltransferase, membrane-bound form
T39368	SLC7A4	Cationic amino acid transporter-4
S74445	RBP5	Cellular retinol-binding protein iii
T55952	RBP5	Cellular retinol-binding protein iii
HSU39905	SLC18A1	Chromaffin granule amine transporter
R52371	SLC35A1	Cmp-sialic acid transporter
D20754	CNT3	Concentrative nucleoside transporter 3
HSMNKMBP	ATP7A	Copper-transporting ATPase 1
HUMWND	ATP7B	Copper-transporting ATPase 2
HUMCFTRM	ABCC7	Cystic fibrosis transmembrane conductance regulator
F10774	SLC7A11	Cystine/glutamate transporter
HUMCYPADA	CYP11B1	Cytochrome P450 11B1, mitochondrial
HUMARM	CYP19	Cytochrome P450 19
HUMCYP145	CYP1A1	Cytochrome P450 1A1
R21282	CYP26	Cytochrome P450 26
AF209774	CYP2A13	Cytochrome P450 2A13
HSC45B2C	CYP2A6	Cytochrome P450 2A6
HSC45B2C	CYP2A7	Cytochrome P450 2A7
HSP452B6	CYP2B6	Cytochrome P450 2B6
HUM2C18	CYP2C18	Cytochrome P450 2C18
HSCP450	CYP2C19	Cytochrome P450 2C19
HUM2C18	CYP2C19	Cytochrome P450 2C19
HUMCYPAX	CYP2C8	Cytochrome P450 2C8
HSCP450	CYP2C9	Cytochrome P450 2C9
HSP450	CYP2D6	Cytochrome P450 2D6
M77918	CYP2E1	Cytochrome P450 2E1
HUMCYPIIF	CYP2F1	Cytochrome P450 2F1
H09076	CYP2J2	Cytochrome P450 2J2
R07010	CYP39A1	Cytochrome P450 39A1
HUMCYPHLP	CYP3A3	Cytochrome P450 3A3
HUMCYPHLP	CYP3A4	Cytochrome P450 3A4
AA416822	CYP3A43	Cytochrome P450 3A43
HUMCYP3A	CYP3A5	Cytochrome P450 3A5
T82801	CYP3A7	Cytochrome P450 3A7
HSCYP4AA	CYP4A11	Cytochrome P450 4A11
S67580	CYP4A11	Cytochrome P450 4A11
HUMCP45IV	CYP4B1	Cytochrome P450 4B1
T98002	CYP4F12	Cytochrome P450 4F12
AA377259	CYP4F2	Cytochrome P450 4F2
AI400898	CYP4F8	Cytochrome P450 4F8
HSU09178	DPYD	Dihydropyrimidine dehydrogenase [nadp+]
W03174	DPYD	Dihydropyrimidine dehydrogenase [nadp+]
HUMFMO1	FMO1	Dimethylaniline monooxygenase [n-oxide forming] 1
HSFLMON2R	FMO2	Dimethylaniline monooxygenase [n-oxide forming] 2
T64494	FMO2	Dimethylaniline monooxygenase [n-oxide forming] 2
T40157	FMO3	Dimethylaniline monooxygenase [n-oxide forming] 3
HSFLMON2R	FMO4	Dimethylaniline monooxygenase [n-oxide forming] 4
D12220	FMO5	Dimethylaniline monooxygenase [n-oxide forming] 5
H25503	HET	Efflux transporter like protein
T12485	HET	Efflux transporter like protein
M78151	EPHX1	Epoxide hydrolase 1
T66884	SLC29A1	Equilibrative nucleoside transporter 1

147

HSHNP36	SLC29A2	Equilibrative nucleoside transporter 2
T08444	SLC1A3	Excitatory amino acid transporter 1
HSU01824	SLC1A2	Excitatory amino acid transporter 2
HSU03506	SLC1A1	Excitatory amino acid transporter 3
F07883	SLC1A6	Excitatory amino acid transporter 4
N39099	SLC1A7	Excitatory amino acid transporter 5
F00548	SLC2A9	Facilitative glucose transporter family member glut9
T95337	SLC27A1	Fatty acid transport protein
Z44099	SLC27A1	Fatty acid transport protein
HUMALBP	FABP4	Fatty acid-binding protein, adipocyte
S67314	FABP3	Fatty acid-binding protein, heart
AW605378	FABP2	Fatty acid-binding protein, intestinal
L25227	SLC19A1	Folate transporter 1
HSI15PGN1	FABP6	Gastrotropin
Z40427	G6PT1	Glucose 5-phosphate transporter
D11793	SLC2A1	Glucose transporter type 1,erythrocyte/brain
N27535	SLC2A10	Glucose transporter type 10
T52633	SLC2A11	Glucose transporter type 11
HUMLGTPA	SLC2A2	Glucose transporter type 2,liver
HUMLGTPA	SLC2A2	Glucose transporter type 2,liver
T07239	SLC2A3	Glucose transporter type 3,brain
HUMIRGT	SLC2A4	Glucose transporter type 4,insulin-responsive.
M62105	SLC2A5	Glucose transporter type 5,small intestine
T59518	SLC2A8	Glucose transporter type 8
HUMLGTH1	GSTA1	Glutathione s-transferase a1
HUMLGTH1	GSTA2	Glutathione s-transferase a2
T98291	GSTA3	Glutathione s-transferase a3-3
Z21581	GSTA4	Glutathione s-transferase a4-4
HSGST4	GSTM1	Glutathione s-transferase mu 1
D31291	GSTM2	Glutathione s-transferase mu 2
HSGST4	GSTM2	Glutathione s-transferase mu 2
T08311	GSTM3	Glutathione s-transferase mu 3
HUMGSTM4B	GSTM4	Glutathione s-transferase mu 4
HUMGSTM5	GSTM5	Glutathione s-transferase mu 5
T05391	GSTP1	Glutathione s-transferase p
AA346312	GSTT1	Glutathione s-transferase theta 1
R08187	GSTT2	Glutathione s-transferase theta 2
Z25318	GSTK1	Glutathione s-transferase, mitochondrial
H03163	SLC37A1	Glycerol-3-phosphate transporter
AA363955	SLC5A7	High affinity choline transporter
HSRRMRNA	SLC7A1	High-affinity cationic amino acid transporter-1
R22196	SLC31A1	High-affinity copper uptake protein 1
AA918012	SLC10A2	Ileal sodium/bile acid transporter
F00840	SLC7A5	Large neutral amino acid transporter small subunit 1
M79133	SLC7A5	Large neutral amino acid transporter small subunit 1
Z38621	SLC7A8	Large neutral amino acids transporter small subunit 2
HUMCARAA	CES1	Liver carboxylesterase
S52379	CES1	Liver carboxylesterase
T55488	SLC21A6	Liver-specific organic anion transporter
W78748	SLC5A4	Low affinity sodium-glucose cotransporter
T54842	SLC7A2	Low-affinity cationic amino acid transporter-2
T87799	ABCA7	Macrophage abc transporter
Z17844	LRP	Major vault protein
Z24885	GSTZ1	Maleylacetoacetate isomerase
T39939	MT1A	Metallothionein-IA
R99207	MT1B	Metallothionein-IB

T39939	MT1E	Metallothionein-IE
D11725	MT1F	Metallothionein-IF
S68949	MT1G	Metallothionein-IG
S68954	MT1G	Metallothionein-IG
HSFMET	MT1H	Metallothionein-IH
S52379	MT2A	Metallothionein-II
M78846	MT3	Metallothionein-III
AA570216	MT1K	Metallothionein-IK
S68954	MT1K	Metallothionein-IK
D11725	MT1L	Metallothionein-IL
HSPP15	MT1L	Metallothionein-IL
HSPP15	MT1R	Metallothionein-IR
NM032935	MT4	Metallothionein-IV
HUMGST	MGST1	Microsomal glutathione s-transferase 1
H59104	MGST2	Microsomal glutathione s-transferase 2
T47062	MGST3	Microsomal glutathione s-transferase 3
SSMPCP	SLC25A3	Mitochondrial phosphate carrier protein
R14814	SULT1A3	Monoamine-sulfating phenol sulfotransferase
HUMARYTRAB	SULT1A3	Monoamine-sulfating phenol sulfotransferase
M62141	SLC16A1	Monocarboxylate transporter 1
H90048	SLC16A6	Monocarboxylate transporter 2
F02520	SLC16A2	Monocarboxylate transporter 3
AI005004	SLC16A8	Monocarboxylate transporter 4
T59354	SLC16A3	Monocarboxylate transporter 5
R22416	SLC16A4	Monocarboxylate transporter 6
T78890	SLC16A5	Monocarboxylate transporter 7
F01173	SLC16A7	Monocarboxylate transporter 8
Z41819	ABCB1	Multidrug resistance protein 1
AL041030	ABCB4	Multidrug resistance protein 3
SATHRMRP	ABCC1	Multidrug resistance-associated protein 1
R00050	ABCC4	Multidrug resistance-associated protein 4
M78673	ABCC5	Multidrug resistance-associated protein 5
R99091	ABCC6	Multidrug resistance-associated protein 6
T69749	ABCC6	Multidrug resistance-associated protein 6
D11495	DIA4	Nad(p)h dehydrogenase [quinone] 1
HUMNRAMP	SLC11A1	Natural resistance-associated macrophage protein 1
Z38360	SLC11A2	Natural resistance-associated macrophage protein 2
HUMASCT1A	SLC1A4	Neutral amino acid transporter a
AW237674	SLC1A5	Neutral amino acid transporter b(0)
M78631	SLC3A1	Neutral and basic amino acid transport protein rbat
HSU08021	NNMT	Nicotinamide n-methyltransferase
T87759	SLC22A4	Novel organic cation transporter 1
Z41935	SLC15A2	Oligopeptide transporter, kidney isoform
HSU21936	SLC15A1	Oligopeptide transporter, small intestine isoform
M62053	OAT1	Organic anion transporter 1
H18607	OAT3	Organic anion transporter 3
R16970	OAT4	Organic anion transporter 4
T39111	SLC21A9	Organic anion transporter b
Z41576	SLC21A11	Organic anion transporter oATP-d
T23657	SLC21A12	Organic anion transporter oATP-e
Z21041	SLC21A14	Organic anion transporting polypeptide 14
H75435	SLC21A8	Organic anion transporting polypeptide 8
HSU77086	SLC22A1	Organic cation transporter 1
HSOCTK	SLC22A2	Organic cation transporter 2
R00207	SLC22A3	Organic cation transporter 3
H30224	ORCTL4	Organic cation transporter like 4

149

H25503	ORCTL2	Organic cation transporter-like 2
Z38659	SLC22A5	Organic cation/carnitine transporter 2
AB010438	ORCTL3	Organic-cation transporter like 3
T95621	ORNT1	Ornithine transporter
AA398593	ORNT2	Ornithine transporter 2
R79412	NTT5	Orphan sodium- and chloride-dependent neurotransmitter transporter ntt5
H82347	NTT73	Orphan sodium- and chloride-dependent neurotransmitter transporter ntt73
Z43484	NTT73	Orphan sodium- and chloride-dependent neurotransmitter transporter ntt73
Z44749	SLC25A17	Peroxisomal membrane protein pmp34
HUMARYLSUL	SULT1A1	Phenol-sulfating phenol sulfotransferase 1
HUMARYLSUL	SULT1A2	Phenol-sulfating phenol sulfotransferase 2
D12243	RBP4	Plasma retinol-binding protein
HUMATPAD	ATP12A	Potassium-transporting ATPase alpha chain 2
Z40030	ATP8A1	Potential phospholipid-transporting ATPase ia
Z40188	FIC1	Potential phospholipid-transporting ATPase ic
T86800	SLC31A2	Probable low-affinity copper uptake protein 2
Z41717	PTGIS	Prostacyclin synthase
S78220	PTGS1	Prostaglandin g/h synthase 1
HUMENDOSYN	PTGS2	Prostaglandin g/h synthase 2
T85296	SLC21A2	Prostaglandin transporter
M62053	SLC22A6	Renal organic anion transport protein 1
HSU26209	SLC13A2	Renal sodium/dicarboxylate cotransporter
Z40774	SLC13A2	Renal sodium/dicarboxylate cotransporter
HSNAPI1	SLC17A1	Renal sodium-dependent phosphate transport protein 1
HUMNAPI3X	SLC34A1	Renal sodium-dependent phosphate transport protein 2
H85361	ABCA4	Retinal-specific ATP-binding cassette transporter
S74445	CRABP1	Retinoic acid-binding protein i, cellular
HUMCRABP	CRABP2	Retinoic acid-binding protein ii, cellular
HUMCRBP	RBP1	Retinol-binding protein i, cellular
S57153	RBP1	Retinol-binding protein i, cellular
T07054	RBP2	Retinol-binding protein ii, cellular
T63266	RBP2	Retinol-binding protein ii, cellular
HUMBGT1R	SLC6A12	Sodium- and chloride-dependent betaine transporter
HUMCRTR	SLC6A8	Sodium- and chloride-dependent creatine transporter 1
R20043	SLC6A13	Sodium- and chloride-dependent gaba transporter 2
S70609	SLC6A9	Sodium- and chloride-dependent glycine transporter 1
AA625644	SLC6A5	Sodium- and chloride-dependent glycine transporter 2
M78677	SLC6A6	Sodium- and chloride-dependent taurine transporter
T10761	SLC4A4	Sodium bicarbonate cotransporter nbc1
AA452802	NBC4	Sodium bicarbonate cotransporter nbc4a
HUMCNC	SLC8A1	Sodium/calcium exchanger 1
R20720	SLC8A2	Sodium/calcium exchanger 2
T07666	SLC8A3	Sodium/calcium exchanger 3
T07666	SLC8A3	Sodium/glucose cotransporter 1
HUMSGLCT	SLC5A2	Sodium/glucose cotransporter 2
S83549	SLC9A2	Sodium/hydrogen exchanger 2
HSU66088	SLC5A5	Sodium/iodide cotransporter
HSU62966	SLC28A1	Sodium/nucleoside cotransporter 1
AA358822	SLC28A2	Sodium/nucleoside cotransporter 2
HUMNTCP	SLC10A1	Sodium/taurocholate cotransporting polypeptide
HSGAT1MR	SLC6A1	Sodium-and chloride-dependent gaba transporter 1
F05686	SLC6A11	Sodium-and chloride-dependent gaba transporter 3
AA604857	SVCT1	Sodium-denpendent vitamin c transporter 1

T27309	SVCT2	Sodium-denpendent vitamin c transporter 2
S44626	SLC6A3	Sodium-dependent dopamine transporter
Z39412	NADC3	Sodium-dependent high-affinity dicarboxylate transporter
T77525	SLC5A6	Sodium-dependent multivitamin transporter
HUMNORTR	SLC6A2	Sodium-dependent noradrenaline transporter
HSZ83953	SLC17A3	Sodium-dependent phosphate transport protein 3
R06460	SLC17A3	Sodium-dependent phosphate transport protein 3
HSZ83953	SLC17A4	Sodium-dependent phosphate transport protein 4
HSY10506	SLC17A4	Sodium-dependent phosphate transport protein 4
H40741	SLC6A7	Sodium-dependent proline transporter
HSSERT	SLC6A4	Sodium-dependent serotonin transporter
T64950	SLC21A3	Sodium-independent organic anion transporter
M79233	EPHX2	Soluble epoxide hydrolase
Z39813	SLC25A18	Solute carrier
HUMSTAR	STAR	Steroidogenic acute regulatory protein
Z20453	STAR	Steroidogenic acute regulatory protein
R69741	SLC26A2	Sulfate transporter
T08860	ABCC8	Sulfonylurea receptor 1
R73927	ABCC9	Sulfonylurea receptor 2
T84623	SULT1C1	Sulfotransferase 1C1
R58632	SULT1C2	Sulfotransferase 1C2
T95810	SLC18A2	Synaptic vesicle amine transporter
AF080246	TRAG3	Taxol resistant associated protein 3
R20880	SLC19A2	Thiamine transporter 1
HSU44128	SLC12A3	Thiazide-sensitive sodium-chloride cotransporter
S62904	TPMT	Thiopurine s-methyltransferase
HSPBX2	G17	Transporter protein
T62038	G17	Transporter protein
R53836	SLC35A3	UDP n-acetylglucosamine transporter
T60594	SLC35A2	UDP-galactose translocator
HUMUGT1FA	UGT1	UDP-glucuronosyltransferase 1-1, microsomal
HUMUGT1FA	UGT1A10	UDP-glucuronosyltransferase 1A10
HUMUGT1FA	UGT1A7	UDP-glucuronosyltransferase 1A7
HUMUGT1FA	UGT1A8	UDP-glucuronosyltransferase 1A8
HUMUGT1FA	UGT1A9	UDP-glucuronosyltransferase 1A9
HSUGT2BIO	UGT2B10	UDP-glucuronosyltransferase 2B10, microsomal
HSUDPGT	UGT2B11	UDP-glucuronosyltransferase 2B11, microsomal
N70316	UGT2B11	UDP-glucuronosyltransferase 2B11, microsomal
HSU08854	UGT2B15	UDP-glucuronosyltransferase 2B15, microsomal
T24450	UGT2B17	UDP-glucuronosyltransferase 2B17, microsomal
HSUDPGT	UGT2B4	UDP-glucuronosyltransferase 2B4, microsomal
HUMUDPGTA	UGT2B7	UDP-glucuronosyltransferase 2B7, microsomal
AI002801	SLC14A1	Urea transporter, erythrocyte
Z19313	SLC14A1	Urea transporter, erythrocyte
AI002801	SLC14A2	Urea transporter, kidney
HSU09210	SLC18A3	Vesicular acetylcholine transporter
HUMKCHB	KCNA4	Voltage-gated potassium channel protein kv1.4
R09608	XDH	Xanthine dehydrogenase/oxidase
T64266	SLC7A7	Y+l amino acid transporter 1
T10628	SLC30A1	Zinc transporter 1
AA322641	SLC30A4	Zinc transporter 4

#DISEASE_RELATED_CLINICAL_PHENOTYPE – This field denotes the possibility of using biomolecular sequences of the present invention for the diagnosis

151

and/or treatment of genetic diseases such as listed in the following URL:
www.geneclinics.org/servlet/access?id=8888891&key=X9D790O5re1Az&db=genete
sts&res=&fcn=b&grp=g&genesearch=true&testtype=both&ls=1&type=e&qry=&sub
mit=Search and in Table 10, below. This list includes genetic diseases and genes
5   which may be used for the detection and/or treatment thereof. As such, newly
uncovered variants of these genes may be used for improved diagnosis and/or
treatment when used singly or in combination with the previously described genes.

Table 10

Gencarta Contig	Gene Symbol	Disease
HSCFTRMA	CFTR	Congenital Bilateral Absence of the Vas Deferens ;Cystic Fibrosis
HUMCFTR M	CFTR	Congenital Bilateral Absence of the Vas Deferens ;Cystic Fibrosis
HUMFGFR3	FGFR3	Achondroplasia ;Crouzon Syndrome with Acanthosis Nigricans ;FGFR-Related Craniosynostosis Syndromes ;Hypochondroplasia ;Muenke Syndrome ;Severe Achondroplasia with Developmental Delay and Acanthosis Nigricans (SADDAN) ;Thanatophoric Dysplasia
T07012	FGD1	Aarskog Syndrome
HSCA1III	COL3A1	Ehlers-Danlos Syndrome, Vascular Type
HUMCOL2 A1B	COL2A1	Achondrogenesis Type 2 ;Kniest Dysplasia ;Spondyloepimetaphyseal Dysplasia, Strudwick Type ;Spondyloepiphyseal Dysplasia, Congenita ;Stickler Syndrome ;Stickler Syndrome Type I
R68817	APRT	Adenine Phosphoribosyltransferase Deficiency
HUMAMPD 1	AMPD1	Adenosine Monophosphate Deaminase 1
M62124	PXR1	Zellweger Syndrome Spectrum
HSXLALDA	ABCD1	Adrenoleukodystrophy, X-Linked
T28718	BTK	X-Linked Agammaglobulinemia
R91110	IL2RG	X-Linked Severe Combined Immunodeficiency
HUMPEDG	OCA2	Oculocutaneous Albinism Type 2
HSU01873	TYR	Oculocutaneous Albinism Type 1
HSOA1MRN A	OA1	Ocular Albinism, X-Linked
R14843	TYRP1	Oculocutaneous Albinism Type 3 (TRP1 Related)
HSALDAR	ALDOA	Aldolase A Deficiency
T40633	HBA1	Alpha-Thalassemia
T40633	HBA2	Alpha-Thalassemia ;Hemoglobin Constant Spring
HSU09820	ATRX	Alpha-Thalassemia X-Linked Mental Retardation Syndrome
HUMCOL4 A5	COL4A5	Alport Syndrome ;Alport Syndrome, X-Linked
T61627	APOE	Apolipoprotein E Genotyping ;Familial Combined Hyperlipidemia ;Hyperlipoproteinemia Type III
T89701	PSEN1	Alzheimer Disease Type 3 ;Early-Onset Familial Alzheimer Disease

R05822	PSEN2	Alzheimer Disease Type 4 ;Early-Onset Familial Alzheimer Disease
HSTTRM	TTR	Transthyretin Amyloidosis
T23978	SOD1	Amyotrophic Lateral Sclerosis
HUMANDREC	AR	Androgen Insensitivity Syndrome ;Spinal and Bulbar Muscular Atrophy
Z19491	UBE3A	Angelman Syndrome
HUMPAX6AN	PAX6	Aniridia ;Anophthalmia ;Isolated Aniridia ;Peters Anomaly ;Peters Anomaly with Cataract ;Wilms Tumor-Aniridia-Genital Anomalies-Retardation Syndrome
HUMKGFRA	FGFR2	Apert Syndrome ;Beare-Stevenson Syndrome ;Crouzon Syndrome ;FGFR-Related Craniosynostosis Syndromes ;Jackson-Weiss Syndrome ;Pfeiffer Syndrome Type 1, 2, and 3
HSU03272	FBN2	Congenital Contractural Arachnodactyly
Z19459	AMCD1	Arthrogryposis Multiplex Congenita, Distal, Type I
T88756	ATM	Ataxia-Telangiectasia
H30056	BBS1	Bardet-Biedl Syndrome
Z25009	BBS2	Bardet-Biedl Syndrome
T64876	BBS4	Bardet-Biedl Syndrome
N27125	PTCH	Nevoid Basal Cell Carcinoma Syndrome
N25339	VMD2	Best Vitelliform Macular Dystrophy
N71795	VMD2	Best Vitelliform Macular Dystrophy
HUMHBB3E	HBB	Beta-Thalassemia ;Hemoglobin E ;Hemoglobin S Beta-Thalassemia ;Hemoglobin SC ;Hemoglobin SD ;Hemoglobin SO ;Hemoglobin SS ;Sickle Cell Disease
H53763	BLM	Bloom Syndrome
N22283	EYA1	Branchiootorenal Syndrome
H90415	BRCA1	BRCA1 and BRCA2 Hereditary Breast/Ovarian Cancer ;BRCA1 Hereditary Breast/Ovarian Cancer
H47777	BRCA2	BRCA1 and BRCA2 Hereditary Breast/Ovarian Cancer ;BRCA2 Hereditary Breast/Ovarian Cancer
Z33575	SOX9	Campomelic Dysplasia
S67156	ASPA	Canavan Disease
T52465	CPS1	Carbamoylphosphate Synthetase I Deficiency
HSVD3HYD	CYP27A1	Cerebrotendinous Xanthomatosis
S66705	MPZ	Charcot-Marie-Tooth Neuropathy Type 1 ;Charcot-Marie-Tooth Neuropathy Type 1B ;Congenital Hypomyelination
HSGAS3MR	PMP22	Charcot-Marie-Tooth Neuropathy Type 1 ;Charcot-Marie-Tooth Neuropathy Type 1A ;Charcot-Marie-Tooth Neuropathy Type 1E ;Hereditary Neuropathy with Liability to Pressure Palsies
T93208	PMP22	Charcot-Marie-Tooth Neuropathy Type 1 ;Charcot-Marie-Tooth Neuropathy Type 1A ;Charcot-Marie-Tooth Neuropathy Type 1E ;Hereditary Neuropathy with Liability to Pressure Palsies
HSGAPJR	GJB1	Charcot-Marie-Tooth Neuropathy Type X
HSXCGD	CYBB	Chronic Granulomatous Disease
S67289	CYBB	Chronic Granulomatous Disease
HSASD	ASS	Citrullinemia
HUMPAX2A	PAX2	Anophthalmia ;Renal-Coloboma Syndrome
HUMP45C21	CYP21A2	21-Hydroxylase Deficiency

S74720	NR0B1	Complex Glycerol Kinase Deficiency ;Dosage-Sensitive Sex Reversal ;Isolated X-Linked Adrenal Hypoplasia Congenita ;X-Linked Adrenal Hypoplasia Congenita
HSKERTRN S	TGM1	Autosomal Recessive Congenital Ichthyosis
BF928311	CPO	Hereditary Coproporphyria
HSCPPOX	CPO	Hereditary Coproporphyria
HUMTGFBI G	TGFBI	Avellino Corneal Dystrophy ;Granular Corneal Dystrophy ;Lattice Corneal Dystrophy Type I
R08437	MSX2	Craniosynostosis Type II ;Parietal Foramina 1
HUMPRP0A	PRNP	Prion Diseases
T08652	DRPLA	DRPLA
Z46151	DRPLA	DRPLA
HSWT1	WT1	Denys-Drash Syndrome ;Wilms Tumor ;Wilms Tumor-Aniridia-Genital Anomalies-Retardation Syndrome ;WT1-Related Disorders
T52050	WT1	Denys-Drash Syndrome ;Wilms Tumor ;Wilms Tumor-Aniridia-Genital Anomalies-Retardation Syndrome ;WT1-Related Disorders
M78080	ATP2A2	Darier Disease
Z30219	DCR	Down Syndrome Critical Region
T11279	DKC1	Dyskeratosis Congenita
T08131	DYT1	Early-Onset Primary Dystonia (DYT1)
T50729	ED1	Hypohidrotic Ectodermal Dysplasia ;Hypohidrotic Ectodermal Dysplasia, X-Linked
HUMPA1V	COL5A1	Ehlers-Danlos Syndrome, Classic Type
HUMLYSY L	PLOD	Ehlers-Danlos Syndrome, Kyphoscoliotic Form
HSCOL1A	COL1A2	Ehlers-Danlos Syndrome, Arthrochalasia Type ;Osteogenesis Imperfecta
HUMCG1PA 1	COL1A1	Ehlers-Danlos Syndrome, Arthrochalasia Type ;Osteogenesis Imperfecta
Z30171	TAZ	3-Methylglutaconic Aciduria Type 2 ;Cardiomyopathy ;Dilated Cardiomyopathy ;Endocardial Fibroelastosis ;Familial Isolated Noncompaction of Left Ventrical Myocardium
Z39302	TAZ	3-Methylglutaconic Aciduria Type 2 ;Cardiomyopathy ;Dilated Cardiomyopathy ;Endocardial Fibroelastosis ;Familial Isolated Noncompaction of Left Ventrical Myocardium
HUMKERK 5A	KRT5	Epidermolysis Bullosa Simplex
R72295	KRT14	Epidermolysis Bullosa Simplex
HUMKTEP2 A	KRT1	Epidermolytic Hyperkeratosis ;Nonepidermolytic Palmoplantar Hyperkeratosis
HUMK10A	KRT10	Epidermolytic Hyperkeratosis
M78482	CHS1	Chediak-Higashi Syndrome
HSTCD1	CHM	Choroideremia
HSAGALAR	GLA	Fabry Disease
T79651	GLA	Fabry Disease
HUMF5A	F5	Factor V Leiden Thrombophilia ;Factor V R2 Mutation Thrombophilia
HUMFXI	F11	Factor XI Deficiency

M79108	APC	Colon Cancer (APC I1307K related) ;Familial Adenomatous Polyposis
T10619	IKBKAP	Familial Dysautonomia
HUMFMR1	FMR1	Fragile X Syndrome
M78417	FMR2	FRAXE Syndrome
R06415	FRDA	Friedreich Ataxia
HSALDOBR	ALDOB	Hereditary Fructose Intolerance
HUMALFU C	FUCA1	Fucosidosis
M85904	FH	Fumarate Hydratase Deficiency
H85361	ABCA4	Age-Related Macular Degeneration ;Retinitis Pigmentosa, Autosomal Recessive ;Stargardt Disease 1
R31596	GALK1	Galactokinase Deficiency
T53762	GALT	Galactosemia
HUMGCB	GBA	Gaucher Disease
T48672	GBA	Gaucher Disease
HSGCRAR	NR3C1	Glucocorticoid Resistance
S58359	G6PD	Glucose-6-Phosphate Dehydrogenase Deficiency
HSGKTS1	GK	Glycerol Kinase Deficiency
HSRNAGLK	GK	Glycerol Kinase Deficiency
U01120	G6PC	Glycogen Storage Disease Type Ia
HUMGAAA	GAA	Glycogen Storage Disease Type II
F00985	AGL	Glycogen Storage Disease Type III
HUMHGBE	GBE1	Glycogen Storage Disease Type IV
HSPHOSR1	PYGM	Glycogen Storage Disease Type V
D12179	PYGL	Glycogen Storage Disease Type VI
HSHMPFK	PFKM	Glycogen Storage Disease Type VII
HUMGLI3A	GLI3	GLI3-Related Disorders ;Greig Cephalopolysyndactyly Syndrome ;Pallister-Hall Syndrome
F09335	ATP2C1	Hailey-Hailey Disease
M62210	CCM1	Angiokeratoma Corporis Diffusum with Arteriovenous Fistulas ;Familial Cerebral Cavernous Malformation
T59431	HFE	HFE- Associated Hereditary Hemochromatosis
HSALK1A	ACVRL1	Hereditary Hemorrhagic Telangiectasia
HUMENDO	ENG	Hereditary Hemorrhagic Telangiectasia
HUMF8C	F8	Hemophilia A
HUMFVIII	F8	Hemophilia A
HUMCFIX	F9	Hemophilia B
HSU03911	MSH2	Hereditary Non-Polyposis Colon Cancer
Z24775	MLH1	Hereditary Non-Polyposis Colon Cancer
HSRETTT	RET	Hirschsprung Disease ;Multiple Endocrine Neoplasia Type 2
HUMSHH	SHH	Holoprosencephaly 3
N81026	TBX5	Holt-Oram Syndrome
M78262	CBS	Homocystinuria
T06035	IDS	Mucopolysaccharidosis Type II
T03828	HD	Huntington Disease
H27612	IDUA	Mucopolysaccharidosis Type I
M62205	GFAP	Alexander Disease
HUMCD40L	TNFSF5	Hyper IgM Syndrome, X-Linked
HUMPTHR OM	F2	Prothrombin G20210A Thrombophilia
T61466	MTHFR	MTHFR Deficiency ;MTHFR Thermolabile Variant

155

HUMSKM1A	SCN4A	Hyperkalemic Periodic Paralysis Type 1 ;Hypokalemic Periodic Paralysis ;Hypokalemic Periodic Paralysis Type 2 ;Myotonia Congenita, Dominant ;Paramyotonia Congenita
HSU09784	CACNA1S	Hypokalemic Periodic Paralysis ;Hypokalemic Periodic Paralysis Type 1 ;Malignant Hyperthermia Susceptibility
HUMLPLA	LPL	Familial Lipoprotein Lipase Deficiency
HUMPEX	PHEX	Hypophosphatemic Rickets, X-Linked Dominant
M78626	STS	Ichthyosis, X-Linked
R56102	IKBKG	Incontinentia Pigmenti
Z39843	IVD	Isovaleric Acidemia
S60085S1	KAL1	Kallmann Syndrome, X-Linked
T55061	KEL	Kell Antigen Genotyping
HUMGALC	GALC	Krabbe Disease
HUMZFPSR EB	ZNF9	Myotonic Dystrophy Type 2
Z19342	KIF1B	Charcot-Marie-Tooth Neuropathy Type 2
T11351	NPC2	Niemann-Pick Disease Type C
Z39096	NDRG1	Charcot-Marie-Tooth Neuropathy Type 4
AA984421	PRX	Charcot-Marie-Tooth Neuropathy Type 4 ;Charcot-Marie-Tooth Neuropathy Type 4F
HUMRETGC	GUCY2D	Leber Congenital Amaurosis
HSU18991	RPE65	Leber Congenital Amaurosis ;Retinitis Pigmentosa, Autosomal Recessive
C16899	MTND6	Leber Hereditary Optic Neuropathy ;Mitochondrial Disorders ;Mitochondrial DNA-Associated Leigh Syndrome and NARP
AA069417	MTND4	Leber Hereditary Optic Neuropathy ;Mitochondrial Disorders ;Mitochondrial DNA-Associated Leigh Syndrome and NARP
HUMCYP3A	MTND4	Leber Hereditary Optic Neuropathy ;Mitochondrial Disorders ;Mitochondrial DNA-Associated Leigh Syndrome and NARP
HSCPHC22	MTND1	Leber Hereditary Optic Neuropathy ;Mitochondrial Disorders ;Mitochondrial DNA-Associated Leigh Syndrome and NARP
HUMHPRT	HPRT1	Lesch-Nyhan Syndrome
HUMLHHC GR	LHCGR	Leydig Cell Hypoplasia/Agenesis ;Male-Limited Precocious Puberty
HSP53	TP53	Li-Fraumeni Syndrome
Z19198	HADHB	Long Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency
M79018	HADHA	Long Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency
R72332	HADHA	Long Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency
W93500	KCNQ1	Atrial Fibrillation ;Jervell and Lange-Nielsen Syndrome ;LQT 1 ;Romano-Ward Syndrome
S62085	OCRL	Lowe Syndrome
T48981	FBN1	Marfan Syndrome
HUMASFB	ARSB	Mucopolysaccharidosis Type VI

M62202	GNAS	Albright Hereditary Osteodystrophy ;McCune-Albright Syndrome ;Osseus Heteroplasia, Progressive
N46342	SACS	ARSACS
T81605	FANCD2	Fanconi Anemia
H47777	FANCD1	Fanconi Anemia
T23877	ACADM	Medium Chain Acyl-Coenzyme A Dehydrogenase Deficiency
AA906866	PARK2	Parkin Type of Juvenile Parkinson Disease
BE140729	GJB4	Erythrokeratodermia Variabilis
HSU26727	CDKN2A	Familial Malignant Melanoma
T47218	SPINK5	Netherton Syndrome
HSMNKMB P	ATP7A	ATP7A-Related Copper Transport Disorders
R37821	SHFM4	Ectrodactyly
Z38987	GSN	Amyloidosis V
HSARYA	ARSA	Chromosome 22q13.3 Deletion Syndrome ;Metachromatic Leukodystrophy
S68531	COL10A1	Metaphyseal Chondrodysplasia, Schmid Type
T59742	CACNA1A	Episodic Ataxia Type 2 ;Familial Hemiplegic Migraine ;Spinocerebellar Ataxia Type 6
HSCP2	HPS3	Hermansky-Pudlak Syndrome ;Hermansky-Pudlak Syndrome 3
R21301	HPS3	Hermansky-Pudlak Syndrome ;Hermansky-Pudlak Syndrome 3
HUMBGAL RP	GLB1	GM1 Gangliosidosis ;Mucopolysaccharidosis Type IVB
HSU12507	KCNJ2	Andersen Syndrome
R28488	MEN1	Multiple Endocrine Neoplasia Type 1
HUMCOMP	COMP	COMP-Related Multiple Epiphyseal Dysplasia ;Multiple Epiphyseal Dysplasia, Dominant ;Pseudoachondroplasia
H30258	COL9A2	Multiple Epiphyseal Dysplasia, Dominant
T48133	EXT1	Hereditary Multiple Exostoses ;Multiple Exostoses, Type I
T06129	EXT2	Hereditary Multiple Exostoses ;Multiple Exostoses, Type II
T05624	LAMA2	Congenital Muscular Dystrophy with Merosin Deficiency
HSDYSTIA	DMD	Duchenne/Becker Muscular Dystrophy ;Dystrophinopathies ;X-Linked Dilated Cardiomyopathy
HSSTA	EMD	Emery-Dreifuss Muscular Dystrophy, X-Linked
HSU20165	BMPR2	Primary Pulmonary Hypertension
M79239	CAPN3	Calpainopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive
HSU34976	SGCG	Gamma-Sarcoglycanopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive ;Sarcoglycanopathies
HUMADHA	SGCA	Alpha-Sarcoglycanopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive ;Sarcoglycanopathies
AI340083	SGCA	Alpha-Sarcoglycanopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive ;Sarcoglycanopathies
Z25374	SGCB	Beta-Sarcoglycanopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive ;Sarcoglycanopathies

N29439	SGCD	Delta-Sarcoglycanopathy ;Dilated Cardiomyopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive ;Sarcoglycanopathies
N56180	CASQ2	Catecholaminergic Ventricular Tachycardia, Autosomal Recessive
T23560	CHRNB2	Nocturnal Frontal Lobe Epilepsy, Autosomal Dominant
HSCHRNA4 4	CHRNA4	Nocturnal Frontal Lobe Epilepsy, Autosomal Dominant
M78654	CHRNA4	Nocturnal Frontal Lobe Epilepsy, Autosomal Dominant
T86329	CDH23	Usher Syndrome Type 1
D11677	PABPN1	Oculopharyngeal Muscular Dystrophy
AW449267	PCDH15	Usher Syndrome Type 1
HUMCLC	CLCN1	Myotonia Congenita, Dominant ;Myotonia Congenita, Recessive
S86455	DMPK	Myotonic Dystrophy Type 1
T70260	MTM1	Myotubular Myopathy, X-Linked
T12579	LMX1B	Nail-Patella Syndrome
HSTRKT1	TPM3	Nemaline Myopathy
HUMTROPC K	TPM3	Nemaline Myopathy
Z19248	NEB	Nemaline Myopathy
AF030626	AVPR2	Nephrogenic Diabetes Insipidus ;Nephrogenic Diabetes Insipidus, X-Linked
AA780862	NPHS1	Congenital Finnish Nephrosis
T08860	ABCC8	ABCC8-Related Hyperinsulinism ;Familial Hyperinsulinism
AA679741	KCNJ11	Familial Hyperinsulinism ;KCNJ11-Related Hyperinsulinism
M77935	NF1	Neurofibromatosis 1
HSMEORPR A	NF2	Neurofibromatosis 2
T08995	CLN3	CLN3-Related Neuronal Ceroid-Lipofuscinosis ;Neuronal Ceroid-Lipofuscinoses
T72120	CLN2	CLN2-Related Neuronal Ceroid-Lipofuscinosis ;Neuronal Ceroid-Lipofuscinoses
T41059	GRHPR	Hyperoxaluria, Primary, Type 2
HUMGCRF C	FCGR3A	Neutrophil Antigen Genotyping
R21657	NPC1	Niemann-Pick Disease Type C ;Niemann-Pick Disease Type C1
M77961	SMPD1	Niemann-Pick Disease Due to Sphingomyelinase Deficiency
T87256	SUOX	Sulfocysteinuria
D79813	SOST	SOST-Related Sclerosing Bone Dysplasias
T94707	MATN3	Multiple Epiphyseal Dysplasia, Dominant
HSCOL9AL	COL9A1	Multiple Epiphyseal Dysplasia, Dominant
S69208	TNNT1	Nemaline Myopathy
Z19459	TPM2	Nemaline Myopathy
D11793	SLC2A1	Glucose Transporter Type 1 Deficiency Syndrome
HSCHRX	NDP	Norrie Disease
T62791	OPA1	Optic Atrophy 1
Z24812	OFD1	Oral-Facial-Digital Syndrome Type I
HUMOTC	OTC	Ornithine Transcarbamylase Deficiency
R66505	MKKS	Bardet-Biedl Syndrome ;McKusick-Kaufman Syndrome
Z19438	CHAC	Choreoacanthocytosis

HUMRDSA	RDS	Patterned Dystrophy of Retinal Pigment Epithelium ;Retinitis Pigmentosa, Autosomal Dominant
Z30072	PLP1	Hereditary Spastic Paraplegia, X-Linked ;PLP-Related Disorders
HSFGR1IG	FGFR1	FGFR-Related Craniosynostosis Syndromes ;Pfeiffer Syndrome Type 1, 2, and 3
HUMPHH	PAH	Phenylalanine Hydroxylase Deficiency
HSKITCR	KIT	Gastrointestinal Stromal Tumor ;Piebaldism
HSGROW1	GH1	Pituitary Dwarfism I
F00079	GHR	Pituitary Dwarfism II
HSPIT1	POU1F1	Pituitary-Specific Transcription Factor Defects (PIT1)
T58874	SDHD	Familial Nonchromaffin Paragangliomas
HUMINTB3	ITGB3	Integrin, Beta 3 ;Platelet Antigen Genotyping
T09245	PKD1	Polycystic Kidney Disease 1, Autosomal Dominant ;Polycystic Kidney Disease, Autosomal Dominant
T55657 .	PKD2	Polycystic Kidney Disease 2, Autosomal Dominant ;Polycystic Kidney Disease, Autosomal Dominant
T77325	PKD2	Polycystic Kidney Disease 2, Autosomal Dominant ;Polycystic Kidney Disease, Autosomal Dominant
W27963	PKD2	Polycystic Kidney Disease 2, Autosomal Dominant ;Polycystic Kidney Disease, Autosomal Dominant
R05352	PKHD1	Polycystic Kidney Disease, Autosomal Recessive
M77871	PCLD	Polycystic Liver Disease
M78097	UROD	Porphyria Cutanea Tarda
HUMPBG	HMBS	Acute Intermittent Porphyria
HUMRODS A	UROS	Congenital Erythropoietic Porphyria
T10891	AGT	Angiotensinogen
T67463	CTSK	Pycnodysostosis
M77954	PDHA1	Pyruvate Dehydrogenase Deficiency, X-linked
Z19400	PHYH	Refsum Disease, Adult
R07476	PEX1	Zellweger Syndrome Spectrum
Z24965	RCA1	Renal Cell Carcinoma
H37900	RHO	Retinitis Pigmentosa, Autosomal Dominant ;Retinitis Pigmentosa, Autosomal Recessive
T24020	RB1	Retinoblastoma
Z44098	RS1	X-Linked Juvenile Retinoschisis
H84683	RS1	X-Linked Juvenile Retinoschisis
HSRH30A	RHCE	Rh C Genotyping ;Rh E Genotyping
S57971	RHCE	Rh C Genotyping ;Rh E Genotyping
AI282496	RHCE	Rh C Genotyping ;Rh E Genotyping
T11224	RHCE	Rh C Genotyping ;Rh E Genotyping
R60192	PEX7	Refsum Disease, Adult ;Rhizomelic Chondrodysplasia Punctata Type 1
HUMMLC1 AA	MLC1	Megalencephalic Leukoencephalopathy with Subcortical Cysts
M79106	MLC1	Megalencephalic Leukoencephalopathy with Subcortical Cysts
T64905	PITX2	Anophthalmia ;Peters Anomaly ;Rieger Syndrome
Z41163	CREBBP	Rubinstein-Taybi Syndrome
HSBHLH	TWIST1	Saethre-Chotzen Syndrome
F00367	EIF2B1	Childhood Ataxia with Central Nervous System Hypomyelination/Vanishing White Matter

Z20030	EIF2B2	Childhood Ataxia with Central Nervous System Hypomyelination/Vanishing White Matter
Z41323	EIF2B3	Childhood Ataxia with Central Nervous System Hypomyelination/Vanishing White Matter
Z17882	EIF2B4	Childhood Ataxia with Central Nervous System Hypomyelination/Vanishing White Matter
R13846	EIF2B5	Childhood Ataxia with Central Nervous System Hypomyelination/Vanishing White Matter ;Cree Leukoencephalopathy
T03917	HEXB	Sandhoff Disease
HUMSRYA	SRY	XX Male Syndrome ;XY Gonadal Dysgenesis
HUMSCAD	ACADS	Short Chain Acyl-CoA Dehydrogenase Deficiency
HSALAS2R	ALAS2	Sideroblastic Anemia, X-Linked
T47846	GPC3	Simpson-Golabi-Behmel Syndrome
T11069	GUSB	Mucopolysaccharidosis Type VII
T08813	SPG3A	Hereditary Spastic Paraplegia, Dominant ;SPG 3
Z21409	SPG3A	Hereditary Spastic Paraplegia, Dominant ;SPG 3
M77964	SPG4	Hereditary Spastic Paraplegia, Dominant ;SPG 4
N36808	SMN1	Spinal Muscular Atrophy
Z38265	SMN1	Spinal Muscular Atrophy
T06490	SCA1	Spinocerebellar Ataxia Type 1
T55469	SCA2	Spinocerebellar Ataxia Type 2
Z41764	SCA2	Spinocerebellar Ataxia Type 2
T61453	MJD	Spinocerebellar Ataxia Type 3
HUMELASF	ELN	Cutis Laxa, Autosomal Dominant ;Supravalvular Aortic Stenosis
T05970	HEXA	Hexosaminidase A Deficiency
M79184	THRB	Thyroid Hormone Resistance
Z20729	TCOF1	Treacher Collins Syndrome
R48739	TRPS1	Trichorhinophalangeal Syndrome Type I
T77655	TSC1	Tuberous Sclerosis 1 ;Tuberous Sclerosis Complex
M78940	TSC2	Tuberous Sclerosis 2 ;Tuberous Sclerosis Complex
HSFAA	FAH	Tyrosinemia Type I
T39510	TBX3	Ulnar-Mammary Syndrome
HUMM7AA	MYO7A	Usher Syndrome Type 1
W22160	USH1C	Usher Syndrome Type 1
T08506	ACADVL	Very Long Chain Acyl-CoA Dehydrogenase Deficiency
HUMHIPLIND	VHL	Von Hippel-Lindau Syndrome
HUMVWF	VWF	Von Willebrand Disease
HSU02368	PAX3	Waardenburg Syndrome Type I
N64051	WRN	Werner Syndrome
HUMWND	ATP7B	Wilson Disease
T40645	WAS	WAS-Related Disorders
HSLAL	LIPA	Wolman Disease
HSASL1	ASL	Argininosuccinicaciduria
HSAGAGENE	AGA	Aspartylglycosaminuria
T88756	ATD	Asphyxiating Thoracic Dystrophy
Z19164	ASAH	Farber Disease
HUMALD	FBP1	Fructose 1,6 Bisphosphatase Deficiency
HSLDHAR	LDHA	Lactate Dehydrogenase Deficiency
M77886	LDHB	Lactate Dehydrogenase Deficiency
HSU13680	LDHC	Lactate Dehydrogenase Deficiency
Z46189	MAN2B1	Alpha-Mannosidosis

M79249	MANBA	Beta-Mannosidosis
H26723	GALNS	Mucopolysaccharidosis Type IVA
H23053	SLC26A4	DFNB 4 ;Enlarged Vestibular Aqueduct Syndrome ;Nonsyndromic Hearing Loss and Deafness, Autosomal Recessive　　;Pendred Syndrome
HSPGK1	PGK1	Phosphoglycerate Kinase Deficiency
HSU08818	MET	Papillary Renal Carcinoma
M79231	PRCC	Papillary Renal Carcinoma
T08200	GNS	Mucopolysaccharidosis Type IIID
HUMNAGB	NAGA	Schindler Disease
T08881	NEU1	Mucolipidosis I
R81783	SLC17A5	Free Sialic Acid Storage Disorders
HUMAUTO NH	MTATP6	Mitochondrial Disorders ;Mitochondrial DNA-Associated Leigh Syndrome and NARP
F09306	SCA7	Spinocerebellar Ataxia Type 7
AF248482	DAZ	Y Chromosome Infertility
HSU21663	DAZ	Y Chromosome Infertility
T47024	JAG1	Alagille Syndrome
HSRYRRM1	RBMY1A1	Y Chromosome Infertility
HSRYRRM2	RBMY1A1	Y Chromosome Infertility
HSVD3R	VDR	Osteoporosis ;Rickets-Alopecia Syndrome
T40157	FMO3	Trimethylaminuria
HUMPHOSL IP	PPGB	Galactosialidosis
HUMPPR	PPGB	Galactosialidosis
H22222	FANCC	Fanconi Anemia
D12009	RPS6KA3	Coffin-Lowry Syndrome
M78282	PTEN	PTEN Hamartoma Tumor Syndrome (PHTS)
M78802	FY	Duffy Antigen Genotyping
HSU04270	KCNH2	LQT 2 ;Romano-Ward Syndrome
T19733	SCN5A	Brugada Syndrome ;LQT 3 ;Romano-Ward Syndrome
HSTFIIDX	TBP	Spinocerebellar Ataxia Type17
HUMKCHA	KCNA1	Episodic Ataxia Type 1
HSU78110	NRTN	Hirschsprung Disease
HSET3AA	EDN3	Hirschsprung Disease
Z17351	ECE1	Hirschsprung Disease
T47284	DHCR7	Smith-Lemli-Opitz Syndrome
HUMXIHB	HBZ	Alpha-Thalassemia
HSCP2	CP	Aceruloplasminemia
N25320	CLN6	CLN6-Related Neuronal Ceroid-Lipofuscinosis ;Neuronal Ceroid-Lipofuscinoses
T11340	NBS1	Nijmegen Breakage Syndrome
Z40114	NBS1	Nijmegen Breakage Syndrome
HSU03688	CYP1B1	Glaucoma, Recessive (Congenital)　　;Peters Anomaly
D62980	MYOC	Glaucoma, Dominant (Juvenile Onset)
T98453	NAGLU	Mucopolysaccharidosis Type IIIB
AA779817	RUNX2	Cleidocranial Dysplasia
HUMCBFA	RUNX2	Cleidocranial Dysplasia
HSMAREN O	MEFV	Familial Mediterranean Fever
F02180	PHKB	Phosphorylase Kinase Deficiency of Liver and Muscle
D11905	HPS1	Hermansky-Pudlak Syndrome ;Hermansky-Pudlak Syndrome 1
R95987	CRX	Retinitis Pigmentosa, Autosomal Dominant
T05762	EVC	Ellis-van Creveld Syndrome

T12126	FLNA	Frontometaphyseal Dysplasia ;Melnick-Needles Syndrome ;Otopalatodigital Syndrome ;Periventricular Heterotopia, X-Linked
T60913	EBP	Chondrodysplasia Punctata, X-Linked Dominant
HSHNF4	HNF4A	Maturity-Onset Diabetes of the Young Type I
HUMBGLU KIN	GCK	Familial Hyperinsulinism ;GCK-Related Hyperinsulinism ;Maturity-Onset Diabetes of the Young Type II
M62026	GCK	Familial Hyperinsulinism ;GCK-Related Hyperinsulinism ;Maturity-Onset Diabetes of the Young Type II
R94860	CIAS1	Chronic Infantile Neurological Cutaneous and Articular Syndrome ;Familial Cold Urticaria ;Muckle-Wells Syndrome
T08221	SMARCAL 1	Schimke Immunoosseous Dysplasia
T95621	SLC25A15	Hyperornithinemia-Hyperammonemia-Homocitrullinuria Syndrome
HUMOATC	OAT	Ornithine Aminotransferase Deficiency
R08989	MLYCD	Malonyl-CoA Decarboxylase Deficiency
N35888	PMM2	Congenital Disorders of Glycosylation
HSRPMI	MPI	Congenital Disorders of Glycosylation
HSSRECV6	MGAT2	Congenital Disorders of Glycosylation
T91755	MGAT2	Congenital Disorders of Glycosylation
HSCPTI	CPT1A	Carnitine Palmitoyltransferase IA (liver) Deficiency
D12096	CPT2	Carnitine Palmitoyltransferase II Deficiency
HSA1ATCA	SERPINA1	Alpha-1-Antitrypsin Deficiency
N36808	SMN2	Spinal Muscular Atrophy
Z38265	SMN2	Spinal Muscular Atrophy
HUMACAD L	ACADL	Long Chain Acyl-CoA Dehydrogenase Deficiency
Z25247	CACT	Carnitine-Acylcarnitine Translocase Deficiency
HUMETFA	ETFA	Glutaricacidemia Type 2
HSETFBS	ETFB	Glutaricacidemia Type 2
S69232	ETFDH	Glutaricacidemia Type 2
T09377	MEB	Muscle-Eye-Brain Disease
Z40427	G6PT1	Glycogen Storage Disease Type Ib
AI002801	SLC14A1	Kidd Genotyping
Z19313	SLC14A1	Kidd Genotyping
HUMPGAM M	PGAM2	Phosphoglycerate Mutase Deficiency
H86930	MPP4	Retinitis Pigmentosa, Autosomal Recessive
HSU14910	RGR	Retinitis Pigmentosa, Autosomal Recessive
AA775466	CARD15	Crohn Disease
AA306952	GAN	Giant Axonal Neuropathy
T99245	CLCN5	Dent Disease
T23537	NR3C2	Pseudohypoaldosteronism Type 1, Dominant
HSLASNA	SCNN1A	Pseudohypoaldosteronism Type 1, Recessive
H26938	SCNN1B	Pseudoaldosteronism ;Pseudohypoaldosteronism Type 1, Recessive
HUMGAMM	SCNN1G	Pseudoaldosteronism ;Pseudohypoaldosteronism Type 1, Recessive
HSP450AL	CYP11B2	Familial Hyperaldosteronism Type 1 ;Familial Hypoaldosteronism Type 2
HUMCYPA DA	CYP11B1	Familial Hyperaldosteronism Type 1
AF017089	COL11A1	Stickler Syndrome ;Stickler Syndrome Type II

HUMCA1XI A	COL11A1	Stickler Syndrome ;Stickler Syndrome Type II
HUMA2XIC OL	COL11A2	Stickler Syndrome
S61523	PIGA	Paroxysmal Nocturnal Hemoglobinuria
T58881	PHKA2	Glycogen Storage Disease Type IX
Z39614	DHAPAT	Rhizomelic Chondrodysplasia Punctata Type 2
N89899	SH2D1A	Lymphoproliferative Disease, X-Linked
HUMUGT1F A	UGT1A1	Gilbert Syndrome
HUMNC1A	COL7A1	Epidermolysis Bullosa Dystrophica, Bart Type ;Epidermolysis Bullosa Dystrophica, Cockayne-Touraine Type ;Epidermolysis Bullosa Dystrophica, Hallopeau-Siemens Type ;Epidermolysis Bullosa Dystrophica, Pasini Type ;Epidermolysis Bullosa, Pretibial
T49684	ITGB4	Epidermolysis Bullosa Letalis with Pyloric Atresia
S66196	ITGA6	Epidermolysis Bullosa Letalis with Pyloric Atresia
T10988	LAMC2	Epidermolysis Bullosa Junctional, Herlitz-Pearson Type
HUMLAMA A	LAMA3	Epidermolysis Bullosa Junctional, Herlitz-Pearson Type
Z24848	LAMA3	Epidermolysis Bullosa Junctional, Herlitz-Pearson Type
T10484	LAMB3	Epidermolysis Bullosa Junctional, Disentis Type ;Epidermolysis Bullosa Junctional, Herlitz-Pearson Type
HUMBP180 AA	COL17A1	Epidermolysis Bullosa Junctional, Disentis Type
M78889	PLEC1	Epidermolysis Bullosa with Muscular Dystrophy
Z38659	SLC22A5	Carnitine Deficiency, Systemic
T85099	CTNS	Cystinosis
W27253	CNGA3	Achromatopsia ;Achromatopsia 2
HSU66088	SLC5A5	Thyroid Hormonogenesis Defect I
HUMTEKRP TK	TEK	Venous Malformation, Multiple Cutaneous and Mucosal
R69741	SLC26A2	Achondrogenesis Type 1B ;Atelosteogenesis Type 2 ;Diastrophic Dysplasia ;Multiple Epiphyseal Dysplasia, Recessive
R70146	PEX10	Zellweger Syndrome Spectrum
S55790	COL4A3	Alport Syndrome ;Alport Syndrome, Autosomal Recessive
HSCOL4A4	COL4A4	Alport Syndrome ;Alport Syndrome, Autosomal Recessive
T10559	SHFM3	Ectrodactyly
T99040	FANCA	Fanconi Anemia
H47777	FANCB	Fanconi Anemia
AA542822	FANCE	Fanconi Anemia
HUMPSPB	PSAP	Metachromatic Leukodystrophy
HUMSAPA1	PSAP	Metachromatic Leukodystrophy
S69686	PSAP	Metachromatic Leukodystrophy
AA252786	NCF1	Chronic Granulomatous Disease
HUMNCF1A	NCF1	Chronic Granulomatous Disease
HSTGFB1	TGFB1	Camurati-Engelmann Disease
R24242	CYBA	Chronic Granulomatous Disease
HUMNOXF	NCF2	Chronic Granulomatous Disease
S41458	PDE6B	Retinitis Pigmentosa, Autosomal Recessive
AA002150	PDE6B	Retinitis Pigmentosa, Autosomal Recessive
R21727	DYSF	Dysferlinopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive
AF055580	USH2A	Usher Syndrome Type 2 ;Usher Syndrome Type 2A

N36632	MITF	Waardenburg Syndrome Type II ;Waardenburg Syndrome Type IIA
M78027	MYH9	DFNA 17 ;Epstein Syndrome ;Fechtner Syndrome ;May-Hegglin Anomaly ;Sebastian Syndrome
Z40194	HPS4	Hermansky-Pudlak Syndrome
AA333774	GP1BA	Platelet Antigen Genotyping
M79110	GP1BB	Platelet Antigen Genotyping
HUMGPIIBA	ITGA2B	Platelet Antigen Genotyping
T29174	ITGA2	Glycoprotein 1a Deficiency ;Platelet Antigen Genotyping
HSGST4	GSTM1	Lung Cancer
AA773443	CHEK2	Li-Fraumeni Syndrome
T78869	CHEK2	Li-Fraumeni Syndrome
T03839	SH3BP2	Cherubism
T67412	IRF6	IRF6-Related Disorders
AB037973	FGF23	Hypophosphatemic Rickets, Dominant
T60199	FBLN5	Cutis Laxa, Autosomal Recessive
T03890	ARX	ARX-Related Disorders
M79175	NSD1	Sotos Syndrome
T07860	NSD1	Sotos Syndrome
M79181	COH1	Cohen Syndrome
MIHS75KDA	NDUFS1	Leigh Syndrome (nuclear DNA mutation) ;Mitochondrial Respiratory Chain Complex I Deficiency
T09312	NDUFV1	Leigh Syndrome (nuclear DNA mutation) ;Mitochondrial Respiratory Chain Complex I Deficiency
AA399371	SALL4	Acrorenoocular Syndrome ;Okihiro Syndrome
HUMA8SEQ	TIMP3	Pseudoinflammatory Fundus Dystrophy
Z40623	GDAP1	Charcot-Marie-Tooth Neuropathy Type 4 ;Charcot-Marie-Tooth Neuropathy Type 4A
AA128030	FOXL2	Blepharophimosis, Epicanthus Inversus, Ptosis
HUMCRTR	SLC6A8	Creatine Deficiency Syndrome, X-Linked
T08882	JPH3	Huntington Disease-Like 2
T07283	SNRPN	Autistic Disorder ;Pervasive Developmental Disorders
Z38837	SPR	Sepiapterin Reductase Deficiency (SR)
HUMANTIR	AGTR1	Angiotensin II Receptor, Type 1
T46961	SEPN1	Congenital Muscular Dystrophy with Early Spine Rigidity ;Multiminicore Disease
Z43954	TRIM32	Limb-Girdle Muscular Dystrophies, Autosomal Recessive
Z19219	TTID	Limb-Girdle Muscular Dystrophies, Autosomal Dominant
HSECADH	CDH1	Hereditary Diffuse Gastric Cancer
Z41199	WFS1	Nonsyndromic Low-Frequency Sensorineural Hearing Loss ;Wolfram Syndrome
HUMLORA	LOR	Progressive Symmetric Erythrokeratoderma
Z38324	HR	Alopecia Universalis ;Papular Atrichia
T09039	RYR1	Central Core Disease of Muscle ;Malignant Hyperthermia Susceptibility ;Multiminicore Disease
T10442	GALE	Galactose Epimerase Deficiency
D82541	PDB2	Paget Disease of Bone
HSU20759	CASR	Autosomal Dominant Hypocalcemia ;Familial Hypocalciuric Hypercalcemia, Type I ;Familial Isolated Hypoparathyroidism ;Neonatal Severe Primary Hyperparathyroidism
AA071082	SALL1	Townes-Brocks Syndrome

T81692	EDAR	Hypohidrotic Ectodermal Dysplasia ;Hypohidrotic Ectodermal Dysplasia, Autosomal
HUMHPA1B	HP	Anhaptoglobinemia
HSU01922	TIMM8A	Deafness-Dystonia-Optic Neuronopathy Syndrome
HUMHSDI	HSD3B2	Prostate Cancer
HSU05659	HSD17B3	Prostate Cancer
Z38915	NPHP4	Nephronophthisis 4 ;Senior-Loken Syndrome
HSC1INHR	SERPING1	Hereditary Angioneurotic Edema
D62739	BBS7	Bardet-Biedl Syndrome
T64266	SLC7A7	Lysinuric Protein Intolerance
S52028	CTH	Cystathioninuria
Z30254	EFEMP1	Doyne Honeycomb Retinal Dystrophy ;Patterned Dystrophy of Retinal Pigment Epithelium
D59254	ELOVL4	Stargardt Disease 3
S43856	GCH1	Dopa-Responsive Dystonia ;GTP Cyclohydrolase 1-Deficient DRD ;GTP Cyclohydrolase-1 Deficiency (GTPCH)
M78468	PAFAH1B1	17-Linked Lissencephaly
M78473	PAFAH1B1	17-Linked Lissencephaly
S51033	MID1	Opitz Syndrome, X-Linked
Z40343	MID1	Opitz Syndrome, X-Linked
HUM6PTHS	PTS	Pyruvoyltetrahydropterin Synthase Deficiency
M62103	CIRH1A	North American Indian Childhood Cirrhosis
HSDHPR	QDPR	Dihydropteridine Reductase Deficiency (DHPR)
T23665	FKRP	Congenital Muscular Dystrophy Type 1C ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive
T60498	LRPPRC	Leigh Syndrome, French-Canadian Type
BG772870	LRPPRC	Leigh Syndrome, French-Canadian Type
HSACHRA	CHRNA1	Congenital Myasthenic Syndromes
HSACHRB	CHRNB1	Congenital Myasthenic Syndromes
HSACHRG	CHRND	Congenital Myasthenic Syndromes
HSACETR	CHRNE	Congenital Myasthenic Syndromes
HSACRAP	RAPSN	Congenital Myasthenic Syndromes
M78334	COLQ	Congenital Myasthenic Syndromes
S56138	CHAT	Congenital Myasthenic Syndromes
D11584	SDHC	Familial Nonchromaffin Paragangliomas
HSPSTI	SPINK1	Hereditary Pancreatitis
HSSPROTR	PROS1	Protein S Heerlen Variant
HUMLAP	ITGB2	Leukocyte Adhesion Deficiency, Type 1
T12572	ADAMTS13	Familial Thrombotic Thrombocytopenia Purpura
HUMCOMIIP	SDHB	Carotid Body Tumors and Multiple Extraadrenal Pheochromocytomas
NM005912	MC4R	Obesity
HUMPAX8A	PAX8	Congenital Hypothyroidism
AA037119	FOXE1	Bamforth-Lazarus Syndrome ;Congenital Hypothyroidism
AV754057	FSHB	Isolated Follicle Stimulating Hormone Deficiency
HUMHOMEOA	PCBD	Pterin-4a Carbinolamine Dehydratase Deficiency (PCD)
HSTHR	TH	Dopa-Responsive Dystonia ;Tyrosine Hydroxylase-Deficient DRD
AA219596	ZIC3	Heterotaxy Syndrome

HSU20324	CSRP3	Dilated Cardiomyopathy
HUMPHLAM	PLN	Dilated Cardiomyopathy
F10219	ALMS1	Alstrom Syndrome
T06612	VCL	Dilated Cardiomyopathy
AF388366	USH3A	Usher Syndrome Type 3
Z40797	SGCE	Myoclonus-Dystonia
T08448	RAB7	Charcot-Marie-Tooth Neuropathy Type 2
D12383	GARS	Charcot-Marie-Tooth Neuropathy Type 2
Z36734	HRPT2	HRPT2-Related Disorders
H19914	EDARADD	Hypohidrotic Ectodermal Dysplasia ;Hypohidrotic Ectodermal Dysplasia, Autosomal
T08852	PPT1	Neuronal Ceroid-Lipofuscinoses ;PPT1-Related Neuronal Ceroid-Lipofuscinosis
HUMDRA	SLC26A3	Familial Chloride Diarrhea
R16324	AGPAT2	Berardinelli-Seip Congenital Lipodystrophy
Z41967	BSCL2	Berardinelli-Seip Congenital Lipodystrophy
W28410	OPN1MW	Blue-Mono-Cone-Monochromatic Type Colorblindness
T27896	OPN1LW	Blue-Mono-Cone-Monochromatic Type Colorblindness
AI469991	PHOX2A	Congenital Fibrosis of Extraocular Muscles
HSFSTHR	FSHR	Premature Ovarian Failure, Autosomal Recessive
HSLPH	LCT	Hypolactasia, Adult Type
Z41000	BCS1L	Gracile Syndrome ;Mitochondrial Respiratory Chain Complex III Deficiency
HSCGJP	GJA1	Oculodentodigital Dysplasia
HSPERFP1	PRF1	Familial Hemophagocytic Lymphohistiocytosis 2
M78112	GLUD1	Familial Hyperinsulinism ;GLUD1-Related Hyperinsulinism
Z39336	GLUD1	Familial Hyperinsulinism ;GLUD1-Related Hyperinsulinism
W79230	RAX	Anophthalmia
AF041339	PITX3	Anophthalmia
AA151708	HESX1	Anophthalmia
HSSOXB	SOX3	Anophthalmia ;Mental Retardation, X-Linked, with Growth Hormone Deficiency
HUMHMGBOX	SOX2	Anophthalmia
HSGM2APA	GM2A	GM2 Activator Deficiency
Z19280	GLC1E	Glaucoma, Dominant (Adult Onset)
T20165	PHF6	Borjeson-Forssman-Lehmann Syndrome
Z40394	CMT4B2	Charcot-Marie-Tooth Neuropathy Type 4
HUMIHH	IHH	Brachydactyly Type A1
HUMCDPK	CDK4	Familial Malignant Melanoma
T39355	SBDS	Shwachman-Diamond Syndrome
HSHMPLK	MPL	Amegakaryocytic Thrombocytopenia, Congenital
Z38860	TRIM37	Mulibrey Nanism
M62027	DTNA	Familial Isolated Noncompaction of Left Ventrical Myocardium
Z39175	DDB2	Xeroderma Pigmentosum
T09329	MUTYH	MYH-Associated Polyposis
HUMAPA	APP	Alzheimer Disease Type 1 ;Early-Onset Familial Alzheimer Disease
M79090	GSS	5-Oxoprolinuria
Z26981	OXCT	3-Oxoacid CoA Transferase
D12046	PMS1	Hereditary Non-Polyposis Colon Cancer
T08186	PMS2	Hereditary Non-Polyposis Colon Cancer
R20984	MSH6	Hereditary Non-Polyposis Colon Cancer

T60457	NDUFS4	Leigh Syndrome (nuclear DNA mutation) ;Mitochondrial Respiratory Chain Complex I Deficiency
D30864	NDUFS8	Leigh Syndrome (nuclear DNA mutation)
M78107	SDHA	Leigh Syndrome (nuclear DNA mutation)
R15290	NDUFS7	Leigh Syndrome (nuclear DNA mutation)
HUMPCBA	PC	Pyruvate Carboxylase Deficiency
R11095	AASS	Hyperlysinemia
T23789	PEX3	Zellweger Syndrome Spectrum
T09086	STK11	Peutz-Jeghers Syndrome
T87335	HAL	Histidinemia
Z19082	ALDH4A1	Hyperprolinemia, Type II
Z25227	MADH4	Juvenile Polyposis Syndrome
M78130	XPB	Xeroderma Pigmentosum
T08987	XPD	Xeroderma Pigmentosum
D81449	XPF	Xeroderma Pigmentosum
HSXPGAA	XPG	Xeroderma Pigmentosum
HSAUHMR	AUH	3-Methylglutaconic Aciduria Type 1
T19530	MMAB	Methylmalonicaciduria
Z40169	MMAA	Methylmalonicaciduria
T93695	BCAT1	Hyperleucine-Isoleucinemia
Z41266	BCAT2	Hyperleucine-Isoleucinemia
HSU03506	SLC1A1	Dicarboxylicaminoaciduria
R88591	PRODH	Hyperprolinemia, Type I
T05380	EPM2A	Progressive Myoclonus Epilepsy, Lafora Type
T27227	FANCF	Fanconi Anemia
H49070	FANCF	Fanconi Anemia
Z41736	FANCG	Fanconi Anemia
R66178	ED4	Ectodermal Dysplasia, Margarita Island Type
L25197	KCNE1	Jervell and Lange-Nielsen Syndrome ;LQT 5 ;Romano-Ward Syndrome
HUMUMOD	UMOD	Familial Nephropathy with Gout ;Medullary Cystic Kidney Disease 2
HSU66583	CRYGD	Cataract, Crystalline Aculeiform
HSPHR	PTHR1	Chondrodysplasia, Blomstrand Type
T97980	MTRR	Homocystinuria-Megaloblastic Anemia
S60710	ADSL	Adenylosuccinase deficiency
Z38216	SLC25A19	Amish Lethal Microcephaly
T35049	SLC25A19	Amish Lethal Microcephaly
T11501	DBH	Dopamine Beta-Hydroxylase Deficiency
H11439	NLGN3	Autistic Disorder ;Pervasive Developmental Disorders
R12551	NLGN4	Autistic Disorder ;Pervasive Developmental Disorders
M78212	ATP1A2	Familial Hemiplegic Migraine
T96957	SPCH1	Severe Speech Delay
AI266171	PHOX2B	Congenital Central Hypoventilation Syndrome
BG723199	DSG4	Localized Autosomal Recessive Hypotrichosis
T46918	HSD11B2	Apparent Mineralocorticoid Excess Syndrome
HUMFERLS	FTL	Hyperferritinemia Cataract Syndrome
HUMCKRASA	KRAS2	Familial Pancreatic Cancer
S39383	PTPN11	LEOPARD Syndrome ;Noonan Syndrome
HUMSTAR	STAR	Cholesterol Desmolase Deficiency
Z20453	STAR	Cholesterol Desmolase Deficiency
HUMVPC	AVP	Neurohypophyseal Diabetes Insipidus
M62144	MECP2	Rett Syndrome
HSCA2VR	COL5A2	Ehlers-Danlos Syndrome, Classic Type

167

HUMGENX	TNXB	Ehlers-Danlos-like Syndrome Due to Tenascin-X Deficiency
R02385	TNXB	Ehlers-Danlos-like Syndrome Due to Tenascin-X Deficiency
T39901	LITAF	Charcot-Marie-Tooth Neuropathy Type 1
AA621310	FOXE3	Anophthalmia
H18132	CFC1	Heterotaxy Syndrome
R36719	EBAF	Heterotaxy Syndrome
HSACTIIRE	ACVR2B	Heterotaxy Syndrome
T52017	CRELD1	Heterotaxy Syndrome
D11851	LMNA	Dilated Cardiomyopathy ;Emery-Dreifuss Muscular Dystrophy, Autosomal Dominant ;Familial Partial Lipodystrophy, Dunnigan Type ;Hutchinson-Gilford Progeria Syndrome ;Limb-Girdle Muscular Dystrophies, Autosomal Dominant ;Mandibuloacral Dysplasia
D12062	DSP	Cardiomyopathy, Dilated, with Woolly Hair and Keratoderma ;Keratosis Palmoplantaris Striata
H99382	MSH3	Hereditary Non-Polyposis Colon Cancer
AW205295	NOG	Multiple Synostoses Syndrome
AA135181	GJB3	Erythrokeratodermia Variabilis
F10278	PEO1	Mitochondrial DNA Deletion Syndromes
M62022	MASS1	Febrile Seizures
HUMQBPCA	UQCRB	Mitochondrial Respiratory Chain Complex III Deficiency
HUMEGR2A	EGR2	Charcot-Marie-Tooth Neuropathy Type 1 ;Charcot-Marie-Tooth Neuropathy Type 1D ;Charcot-Marie-Tooth Neuropathy Type 4 ;Charcot-Marie-Tooth Neuropathy Type 4E
HSFLT4X	FLT4	Milroy Congenital Lymphedema
Z24968	PEX26	Zellweger Syndrome Spectrum
AA338362	ANKH	Craniometaphyseal Dysplasia, Dominant
HUMRPS24A	RPS19	Diamond-Blackfan Anemia
T11633	RPS19	Diamond-Blackfan Anemia
HSACMHCP	MYH7	Dilated Cardiomyopathy ;Familial Hypertrophic Cardiomyopathy
Z25920	TNNT2	Dilated Cardiomyopathy ;Familial Hypertrophic Cardiomyopathy
HUMTRO	TPM1	Dilated Cardiomyopathy ;Familial Hypertrophic Cardiomyopathy
Z18303	MYBPC3	Dilated Cardiomyopathy ;Familial Hypertrophic Cardiomyopathy
HSU09466	COX10	Leigh Syndrome (nuclear DNA mutation)
S72487	ECGF1	Mitochondrial Neurogastrointestinal Encephalopathy Syndrome
M62196	KIF5A	Hereditary Spastic Paraplegia, Dominant
T07578	KIF5A	Hereditary Spastic Paraplegia, Dominant
D11648	HSPD1	Hereditary Spastic Paraplegia, Dominant
T47330	SOX18	Hypotrichosis-Lymphedema-Telangiectasia Syndrome
AA448334	CAV3	Caveolinopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Dominant
AW071529	ALX4	Parietal Foramina 2
M61973	CD2AP	Focal Segmental Glomerulosclerosis
W21801	NR2E3	Enhanced S-Cone Syndrome

Z20305	TREM2	PLOSL
T05421	ANK2	LQT 4 ;Romano-Ward Syndrome
HUMROR2 A	ROR2	ROR2-Related Disorders
Z25920	CMD1D	Dilated Cardiomyopathy
AA887962	HLXB9	Currarino Syndrome
R00281	ALDH5A1	Succinic Semialdehyde Dehydrogenase Deficiency
HSPCCAR	PCCA	Propionic Acidemia
N43992	DLL3	Spondylocostal Dysostosis, Autosomal Recessive ;Syndactyly, Type IV
Z39790	MUT	Methylmalonicaciduria
HUMARGL	ARG1	Argininemia
M78631	SLC3A1	Cystinuria
T80665	SLC7A9	Cystinuria
T27286	HGD	Alkaptonuria
HUMBCKD H	BCKDHA	Maple Syrup Urine Disease
HUMBCKD HA	BCKDHB	Maple Syrup Urine Disease
HSTRANSP	DBT	Maple Syrup Urine Disease
Z44722	HLCS	Holocarboxylase Synthetase Deficiency
Z38396	BTD	Biotinidase Deficiency
T48178	POMT1	Walker-Warburg Syndrome
T28737	GJB2	DFNA 3 Nonsyndromic Hearing Loss and Deafness ;DFNB 1 Nonsyndromic Hearing Loss and Deafness ;GJB2-Related DFNA 3 Nonsyndromic Hearing Loss and Deafness ;GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness ;Nonsyndromic Hearing Loss and Deafness, Autosomal Dominant ;Nonsyndromic Hearing Loss and Deafness, Autosomal Recessive ;Vohwinkel Syndrome
T05861	COCH	DFNA 9 (COCH) ;Nonsyndromic Hearing Loss and Deafness, Autosomal Dominant
HSBRN4	POU3F4	DFN 3
HSU21938	TTPA	Ataxia with Vitamin E Deficiency (AVED)
T93783	KIAA1985	Charcot-Marie-Tooth Neuropathy Type 4
BE735997	SANS	Usher Syndrome Type 1
AA548783	HOXD13	Syndactyly, Type II
R33750	HOXA13	Hand-Foot-Uterus Syndrome
HUMPP	GLDC	GLDC-Related Glycine Encephalopathy ;Glycine Encephalopathy
F04230	AMT	AMT-Related Glycine Encephalopathy ;Glycine Encephalopathy
T54795	DECR	2,4-Dienoyl-CoA Reductase Deficiency
R07295	ACAT1	Ketothiolase Deficiency
S70578	ACAT1	Ketothiolase Deficiency
HUMMEVK IN	MVK	Hyper IgD Syndrome ;Mevalonicaciduria
T11245	HMGCL	3-Hydroxy-3-Methylglutaryl-Coenzyme A Lyase Deficiency
Z41427	GCDH	Glutaricacidemia Type 1
HSSHOXA	SHOX	Langer Mesomelic Dwarfism ;Leri-Weill Dyschondrosteosis ;Short Stature
HUMDOPA DC	DDC	Aromatic L-Amino Acid Decarboxylase Deficiency

HSCOL3A4	COL6A3	Limb-Girdle Muscular Dystrophies, Autosomal Dominant
HSCOL1A4	COL6A1	Limb-Girdle Muscular Dystrophies, Autosomal Dominant
HSCOL2C2	COL6A2	Limb-Girdle Muscular Dystrophies, Autosomal Dominant
H16770	RECQL4	Rothmund-Thomson Syndrome
H11473	SGSH	Mucopolysaccharidosis Type IIIA·
H67137	MCCC1	3-Methylcrotonyl-CoA Carboxylase Deficiency
R88931	MCCC2	3-Methylcrotonyl-CoA Carboxylase Deficiency
Z24865	TCAP	Dilated Cardiomyopathy ;Limb-Girdle Muscular Dystrophies, Autosomal Recessive
M86030	DCX	DCX-Related Malformations
HUMACTA SK	ACTA1	Nemaline Myopathy
HSDGIGLY	DSG1	Keratosis Palmoplantaris Striata
HSRETSA	SAG	Retinitis Pigmentosa, Autosomal Recessive
HSAPHOL	ALPL	Hypophosphatasia
N73784	XPA	Xeroderma Pigmentosum
T28958	XPC	Xeroderma Pigmentosum
N69543	POLH	Xeroderma Pigmentosum
T54103	POLH	Xeroderma Pigmentosum
H56484	CKN1	Cockayne Syndrome
Z38185	ERCC6	Cockayne Syndrome
F07041	PI12	Familial Encephalopathy with Neuroserpin Inclusion Bodies
AA633404	KCNE2	LQT 6 ;Romano-Ward Syndrome
AF302095	KCNE2	LQT 6 ;Romano-Ward Syndrome
HSTITINC2	CMD1G	Dilated Cardiomyopathy
N99115	NPHP1	Nephronophthisis 1 ;Senior-Loken Syndrome
HUMELAN AA	ELA2	ELA2-Related Neutropenia
S67325	PCCB	Propionic Acidemia
HSGA7331	M1S1	Corneal Dystrophy, Gelatinous Drop-Like
HSACE	ACE	Angiotensin I Converting Enzyme 1
S49816	TSHR	Congenital Hypothyroidism ;Familial Non-Autoimmune Hyperthyroidism
Z30221	VMGLOM	Multiple Glomus Tumors
H88042	COL9A3	Multiple Epiphyseal Dysplasia, Dominant
M78119	ADA	Adenosine Deaminase Deficiency
T55785	GAMT	Guanidinoacetate Methyltransferase Deficiency
HUMCST4B A	CSTB	Myoclonic Epilepsy of Unverricht and Lundborg
S73196	AQP2	Nephrogenic Diabetes Insipidus ;Nephrogenic Diabetes Insipidus, Autosomal
HSU76388	NR5A1	XY Sex Reversal with Adrenal Failure
HSCPHC22	MTRNR1	MTRNR1-Related Hearing Loss and Deafness
H21596	PPARG	Diabetes Mellitus with Acanthosis Nigricans and Hypertension
D56550	FOXC1	Anophthalmia ;Rieger Syndrome
M78868	AP3B1	Hermansky-Pudlak Syndrome
T47068	NOTCH3	CADASIL
HSHMF1C	TCF1	Maturity-Onset Diabetes of the Young Type III
AA223508	TCF1	Maturity-Onset Diabetes of the Young Type III
AF049893	IPF1	Maturity-Onset Diabetes of the Young Type IV
HSU30329	IPF1	Maturity-Onset Diabetes of the Young Type IV
HSVHNF1	TCF2	Maturity-Onset Diabetes of the Young Type V
HUMLDLRF	LDLR	Familial Hypercholesterolemia

170

MT		
HSAPOBR2	APOB	Familial Hypercholesterolemia Type B
T78010	ABCB7	Sideroblastic Anemia and Ataxia
AF076215	PROP1	PROP1-Related Combined Pituitary Hormone Deficiency
S99468	ALAD	Acute Hepatic Porphyria
T61818	ABCC2	Dubin-Johnson Syndrome
HUMLCAT	LCAT	Lecithin Cholesterol Acyltransferase Deficiency
Z38510	HADHSC	Short Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency, Liver
AF041240	PPOX	Variegate Porphyria
T77011	PPOX	Variegate Porphyria
Z40014	ALDH10	Sjogren-Larsson Syndrome
S79867	KRT16	Nonepidermolytic Palmoplantar Hyperkeratosis ;Pachyonychia Congenita
HUMKER56K	KRT6A	Pachyonychia Congenita
HSKERELP	KRT17	Pachyonychia Congenita ;Steatocystoma Multiplex
R11850	KRT6B	Pachyonychia Congenita
S69510	KRT9	Epidermolytic Palmoplantar Keratoderma
HSCYTK	KRT13	White Sponge Nevus of Cannon
T92918	KRT4	White Sponge Nevus of Cannon
S54769	SPG7	Hereditary Spastic Paraplegia, Recessive ;SPG 7
T50707	FECH	Erythropoietic Protoporphyria
HUMPOMM	PXMP3	Zellweger Syndrome Spectrum
R05392	PEX6	Zellweger Syndrome Spectrum
Z38759	PEX12	Zellweger Syndrome Spectrum
R14480	PEX16	Zellweger Syndrome Spectrum
R10031	PEX13	Zellweger Syndrome Spectrum
R13532	PXF	Zellweger Syndrome Spectrum
Z30136	AGPS	Rhizomelic Chondrodysplasia Punctata Type 3
HSU07866	ACOX	Pseudoneonatal Adrenoleukodystrophy
N63143	ALG6	Congenital Disorders of Glycosylation
HSTNFR1A	TNFRSF1A	Familial Hibernian Fever
AA018811	RP1	Retinitis Pigmentosa, Autosomal Dominant
HSG11	RP1	Retinitis Pigmentosa, Autosomal Dominant
T07942	RP1	Retinitis Pigmentosa, Autosomal Dominant
H28658	PRPF31	Retinitis Pigmentosa, Autosomal Dominant
T07062	PRPF8	Retinitis Pigmentosa, Autosomal Dominant
T05573	RP18	Retinitis Pigmentosa, Autosomal Dominant
HUMNRLGP	NRL	Retinitis Pigmentosa, Autosomal Dominant
T87786	CRB1	Retinitis Pigmentosa, Autosomal Recessive
H92408	TULP1	Retinitis Pigmentosa, Autosomal Recessive
S42457	CNGA1	Retinitis Pigmentosa, Autosomal Recessive
H30568	PDE6A	Retinitis Pigmentosa, Autosomal Recessive
M78192	RLBP1	Retinitis Pigmentosa, Autosomal Recessive ;Retinitis Pigmentosa, Autosomal Recessive, Bothnia Type
T10761	SLC4A4	Proximal Renal Tubular Acidosis with Ocular Abnormalities

N64339	GJB6	DFNA 3 Nonsyndromic Hearing Loss and Deafness ;DFNB 1 Nonsyndromic Hearing Loss and Deafness ;GJB6-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness ;GJB6-Related DFNA 3 Nonsyndromic Hearing Loss and Deafness ;Hidrotic Ectodermal Dysplasia 2 ;Nonsyndromic Hearing Loss and Deafness, Autosomal Dominant ;Nonsyndromic Hearing Loss and Deafness, Autosomal Recessive
T67968	MAT1A	Isolated Persistent Hypermethioninemia
HUMUMPS	UMPS	Oroticaciduria
HSPNP	NP	Purine Nucleoside Phosphorylase Deficiency
AB006682	AIRE	Autoimmune Polyendocrinopathy Syndrome Type 1
BE871354	JUP	Naxos Disease
T08214	JUP	Naxos Disease
F00120	DES	Dilated Cardiomyopathy
R28506	MOCS1	Molybdenum Cofactor Deficiency
T70309	MOCS2	Molybdenum Cofactor Deficiency
T08212	SNCA	Parkinson Disease
R99091	ABCC6	Pseudoxanthoma Elasticum
T69749	ABCC6	Pseudoxanthoma Elasticum
AA207040	PRG4	Arthropathy Camptodactyly Syndrome
T57014	PRG4	Arthropathy Camptodactyly Syndrome
F07016	OPPG	Osteoporosis Pseudoglioma Syndrome
H27782	SCO2	Fatal Infantile Cardioencephalopathy due to COX Deficiency
S54705S1	PRKAR1A	Carney Complex
Z25903	SCA10	Spinocerebellar Ataxia Type10
AA592984	WISP3	Progressive Pseudorheumatoid Arthropathy of Childhood
Z39666	MCOLN1	Mucolipidosis IV
HSEMX2	EMX2	Familial Schizencephaly
HUMSP18A	SFTPB	Pulmonary Surfactant Protein B Deficiency
Z40188	ATP8B1	Benign Recurrent Intrahepatic Cholestasis ;Progressive Familial Intrahepatic Cholestasis ;Progressive Familial Intrahepatic Cholestasis 1
U46845	CYP27B1	Pseudovitamin D Deficiency Rickets
Z21585	MAPT	Frontotemporal Dementia with Parkinsonism-17
HSPPD	HPD	Tyrosinemia Type III
HUMUGT1F A	UGT1A	Crigler-Najjar Syndrome
R20880	SLC19A2	Thiamine-Responsive Megaloblastic Anemia Syndrome
H42203	TFAP2B	Char Syndrome
Z30126	RYR2	Catecholaminergic Ventricular Tachycardia, Autosomal Dominant
HSSPYRAT	AGXT	Hyperoxaluria, Primary, Type 1
T80758	SEDL	Spondyloepiphyseal Dysplasia Tarda, X-Linked
T89449	SEDL	Spondyloepiphyseal Dysplasia Tarda, X-Linked
AA373083	FOXC2	Lymphedema with Distichiasis
HUMPROP2 AB	SCA12	Spinocerebellar Ataxia Type12
Z30145	ACTC	Dilated Cardiomyopathy
HS1900	GDNF	Hirschsprung Disease
M62223	NEFL	Charcot-Marie-Tooth Neuropathy Type 1F/2E ;Charcot-Marie-Tooth Neuropathy Type 2 ;Charcot-Marie-Tooth Neuropathy Type 2E/1F
T10920	SERPINE1	Plasminogen Activator Inhibitor I

HSNCAML1	L1CAM	Hereditary Spastic Paraplegia, X-Linked ;L1 Syndrome
T11074	L1CAM	Hereditary Spastic Paraplegia, X-Linked ;L1 Syndrome
HUMHPROT	GCSH	Glycine Encephalopathy
HSTATR	TAT	Tyrosinemia Type II
Z19514	CPT1B	Carnitine Palmitoyltransferase IB (muscle) Deficiency
BE149388	CPT1B	Carnitine Palmitoyltransferase IB (muscle) Deficiency
HSALK3A	BMPR1A	Juvenile Polyposis Syndrome
T78581	CLN5	CLN5-Related Neuronal Ceroid-Lipofuscinosis ;Neuronal Ceroid-Lipofuscinoses
N32269	CLN8	CLN8-Related Neuronal Ceroid-Lipofuscinosis ;Neuronal Ceroid-Lipofuscinoses
HSU44128	SLC12A3	Gitelman Syndrome
AI590292	NPHS2	Focal Segmental Glomerulosclerosis ;Steroid-Resistant Nephrotic Syndrome
M62209	ACTN4	Focal Segmental Glomerulosclerosis
H53423	CNGB3	Achromatopsia ;Achromatopsia 3
HSEPAR	HCI	Hemangioma, Hereditary
R14741	ZIC2	Holoprosencephaly 5
H84264	SIX3	Anophthalmia ;Holoprosencephaly 2
T10497	TGIF	Holoprosencephaly 4
Z30052	USP9Y	Y Chromosome Infertility
N85185	DBY	Y Chromosome Infertility
T11164	SPTLC1	Hereditary Sensory Neuropathy Type I
T68440	GNE	GNE-Related Myopathies ;Sialuria, French Type
HSPROPERD	PFC	Properdin Deficiency, X-Linked
T46865	SURF1	Leigh Syndrome (nuclear DNA mutation)
AI015025	VAX1	Anophthalmia
BM727523	VAX1	Anophthalmia
AA310724	SIX6	Anophthalmia
R37821	TP63	TP63-Related Disorders
AF091582	ABCB11	Progressive Familial Intrahepatic Cholestasis
HUMHOX7	MSX1	Hypodontia, Autosomal Dominant ;Tooth-and-Nail Syndrome
R15034	CACNB4	Episodic Ataxia Type 2
T52100	TYROBP	PLOSL
F09012	MTMR2	Charcot-Marie-Tooth Neuropathy Type 4
T08510	APTX	Ataxia with Oculomotor Apraxia ;Ataxia with Oculomotor Apraxia 1
HUMHAAC	HF1	Hemolytic-Uremic Syndrome
C16899	MTND5	Leber Hereditary Optic Neuropathy ;Mitochondrial DNA-Associated Leigh Syndrome and NARP

#AUTOANTIGEN_IN_AUTOIMMUNE_DISEASE -Secreted splice variants of known autoantigens associated with a specific autoimmune syndrome, as for example, these listed in table 11, can be used to treat the syndrome. The proposed therapeutic mechanism is that the secreted splice variant would bind the auto-antibodies which formed against the autoantigen, therefore reduce their circulating levels, that would lead

173

to less binding of the autoantigen by auto antibodies and as a consequence diminish the autoimmune clinical symptoms.

Examples of proteins which are involved in autoimmune diseases are presented in Table 11 together with the corresponding internal gene contig name, enabling to allocate 5 the new splice variants within the data files in the attached CD-ROMs 1 and 3.

Table 11

Contig	Disease	Description
HUMROSSA	Sjogren's syndrome	52 kDa Ro protein
HUMI69KAA	Insulin dependent diabetes Mellitus	69 kDa islet cell autoantigen
S55790	Goodpasture's syndrome	alpha 3 chain of collagen IV
HSACHRA	Myasthenia Gravis	Alpha chain of nicotinic Acetyl Choline receptor
Z21711	Rheumatoid Arthritis	Annexin A11
Z21711	Sjogren's syndrome	Annexin A11
Z21711	SLE	Annexin A11
S38729	SLE	ATP-dependent DNA helicase II, 70 kDa subunit
M77907	SLE	ATP-dependent DNA helicase II, 80 kDa subunit
T08224	scleroderma	Autoantigen p27
T08224	Sjogren's syndrome	Autoantigen p27
M85815	Pemphigus	bullous pemphigoid antigen 1
HUMROSSAA	SLE	calreticulin
HUMCENPRO	General autoimmune response	Centromere autoantigen C
HSU14518	General autoimmune response	Centromere protein A
M62116	dermatomyositis	Chromodomain helicase-DNA-binding protein 3
T05980	dermatomyositis	Chromodomain helicase-DNA-binding protein 4
H18687	Autoimmune demyelinating disease	claudin 11
M79258	dermatomyositis	Dermatomyositis associated with cancer putative autoantigen-1
HSDGIGLY	Pemphigus foliaceus	Desmoglein 1
HUMPVA	Pemphigus vulgaris	Desmoglein 3
BG723199	Pemphigus vulgaris	desmoglein 4
M77924	Primary billiary cirrhosis	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial
D11598	Polymyositis	Exosome complex exonuclease RRP45
D11598	scleroderma	Exosome complex exonuclease RRP45
HUMACTINBI	Grave's disease	Filamin B
Z17837	Rheumatoid Arthritis	follistatin-like 1
HUMGAD	Insulin dependent diabetes Mellitus	glutamate decarboxylase 1 (GAD 1)
HSGLAD2A	Insulin dependent diabetes Mellitus	glutamate decarboxylase 2 (GAD 2)
D12383	dermatomyositis	glycyl-tRNA synthetase
D12383	Polymyositis	glycyl-tRNA synthetase
Z40013	Sjogren's syndrome	Golgi autoantigen, golgin subfamily A member 1
N28220	Rheumatoid Arthritis	Golgi autoantigen, golgin subfamily B member 1

174

N28220	Sjogren's syndrome	Golgi autoantigen, golgin subfamily B member 1
HUMMSCA	Grave's disease	Grave's disease carrier protein
HUMGRAVIN	Myasthenia Gravis	gravin
HUMRNPSMBA	SLE	Homo sapiens small nuclear ribonucleoprotein polypeptides B and B1
HUMINSR	Insulin resistant diabetes Mellitus	insulin receptor
HSRNAIFMH	Pernicious Anemia	intrinsic factor
D12018	dermatomyositis	isoleucine-tRNA synthetase
D12018	Polymyositis	isoleucine-tRNA synthetase
T97710	Pemphigus	ladinin 1
HSAUTAN64	Autoimmune thyroid disease	Leiomodin 1
HSLAANT	SLE	Lupus La protein
HUM60RO	SLE	Lupus Ro Protein
F02808	dermatomyositis	lysyl-tRNA synthetase
F02808	Polymyositis	lysyl-tRNA synthetase
F01282	General autoimmune response	Major centromere autoantigen B
M78010	multiple sclerosis	myelin basic protein
R89508	Autoimmune demyelinating disease	Myelin oligodendrocyte glycoprotein (MOG)
HUMHSTNBP	Autoimmune infertility	Nuclear autoantigenic sperm protein
S80305	Antiphospholipid syndrome	Phospholipid beta 2 glycoprotein 1 complex
D11598	Polymyositis	polymyositis/scleroderma autoantigen 1
D11598	scleroderma	polymyositis/scleroderma autoantigen 1
HUMAUA	Polymyositis	Polymyositis/scleroderma autoantigen 2
HUMAUA	scleroderma	Polymyositis/scleroderma autoantigen 2
HUMMCH	Vitiligo	Pro-melanin-concentrating hormone
T05361	Insulin dependent diabetes Mellitus	protein tyrosine phosphatase
HSP3MY	Wegener's granulomatosis	Proteinase 3 (ANCA - antineutrophil cytoplasmic antibody)
F02560	Insulin dependent diabetes Mellitus	Protein-tyrosine phosphatase-like N [Precursor]
T05361	Insulin dependent diabetes Mellitus	Receptor-type protein-tyrosine phosphatase N2
HUM60RO	Sjogren's syndrome	Sjogren syndrome antigen A2
H81770	Sjogren's syndrome	Sjogren's syndrome nuclear autoantigen 1
HUMSNRNPD	SLE	Small nuclear ribonucleoprotein Sm D1
HUMMSCA	Grave's disease	solute carrier family 25
Z17347	Insulin dependent diabetes Mellitus	SOX-13 protein
N79953	Autoimmune infertility	Sperm surface protein Sp17
T08224	scleroderma	SSSCA1
T08224	Sjogren's syndrome	SSSCA1
R54783	interstitial cystitis	synaptonemal complex protein SC65 (SC65)
S40807	Hashimoto's thyroditis	thyroglobulin
S38729	Autoimmune thyroid disease	thyroid autoantigen 70kDa
HUMTPOA	Hashimoto's thyroditis	Thyroid peroxidase
HUMBF7A	Celiac disease	transglutaminase 2
S49816	Grave's disease	TSH receptor

Differentially expressed biomolecular sequences - field description

#TS – This field denotes tissue-specific genes which gene products are upregulated in at least one tissue. Such gene products might be used as tissue or pathological markers. Therapeutic uses of such gene products vary and may include, for example, anti-cancer vaccination and drug-targeting. Other exemplary uses are described hereinabove. It will be appreciated that avary differentially expressed gene product can be assigned to higher hierarchies of classification. Thus, for example, a prostate cancer specific gene product may be used as a diagnostic marker for this cancer, but may be also used as epithelial cancer marker and as a general cancer marker. See for example, Table 12, below.

Table 12

Tissue-tumor searched	Cancer sub-type	Cancer type	Cancer - general
All tumor types			All tumor types
prostate-tumor	prostate-tumor	All epithelial tumors	All tumor types
lung-tumor	lung-tumor	All epithelial tumors	All tumor types
head and neck-tumor	head and neck-tumor	All epithelial tumors	All tumor types
stomach-tumor	stomach-tumor	All epithelial tumors	All tumor types
colon-tumor	colon-tumor	All epithelial tumors	All tumor types
mammary-tumor	mammary-tumor	All epithelial tumors	All tumor types
kidney-tumor	kidney-tumor	All epithelial tumors	All tumor types
ovary-tumor	ovary-tumor	All epithelial tumors	All tumor types
uterus/cervix-tumor	uterus/cervix-tumor	All epithelial tumors	All tumor types
thyroid-tumor	thyroid-tumor	All epithelial tumors	All tumor types
adrenal-tumor	adrenal-tumor	All epithelial tumors	All tumor types
pancreas-tumor	pancreas-tumor	All epithelial tumors	All tumor types
liver-tumor	liver-tumor	All epithelial tumors	All tumor types
skin-tumor	skin-tumor	All epithelial tumors	All tumor types
brain-tumor	brain-tumor		All tumor types
eye-tumor	eye-tumor		All tumor types
bone-tumor	bone-tumor	Sarcoma	All tumor types
bone marrow-tumor	bone marrow-tumor	Blood cancer	All tumor types
blood-cancer	blood-cancer	Blood cacner	All tumor types
T-cells-tumor	T-cells-tumor	Blood cancer	All tumor types
lymph nodes-tumor	lymph nodes-tumor	Blood cancer	All tumor types
muscle-tumor	muscle-tumor	Sarcoma	All tumor types
testis-tumor	testis-tumor		All tumor types

The annotation format of differentially expressed gene products is as follows.

#TS tissue-name – where the "tissue name" field specifies the list of tissues for which tissue-specific genes/variants were searched, as follows: amniotic+placenta; Blood; Bone; Bone marrow; Brain; Cervix+uterus; Colon; Endocrine, adrenal gland; Endocrine, pancreas; Endocrine, parathyroid+thyroid; Gastrointestinal tract;

176

Genitourinary; Head and neck; Immune, T-cells; Kidney; Liver; Lung; Lymph node; Mammary gland; Muscle; Ovary; Prostate; Skin; Thymus.

#TAA – This field denotes genes or transcript sequences over-expressed in cancer. The annotation format is as follows.

5   #TAA tissue-name - - where the "tissue name" field specifies the list of tissues for which tissue-tumor specific genes/variants were searched, as follows: All tumor types; All epithelial tumors; prostate-tumor; lung-tumor; head and neck-tumor; stomach-tumor; colon-tumor; mammary-tumor; kidney-tumor; ovary-tumor; uterus/cervix-tumor; thyroid-tumor; adrenal-tumor; pancreas-tumor; liver-tumor;

10   skin-tumor; brain-tumor; bone-tumor; bone marrow-tumor; blood-cancer; T-cells-tumor; lymph nodes-tumor; muscle-tumor.

EXAMPLE 11

Modulating Nicastrin activity using COPA inhibitors

15   The present inventors uncovered that coatomer protein complex, subunit α (COPA, GenBank Accession No. NM_004371) and Nicastrin (NCSTN, GenBank Accession No. NM_015331) are a 5'-5' antisense pair. Sequence alignment is illustrated in Figure 24a. Expression patterns of NCSTN and COPA show positive correlation (see Figure 24b).

20   Nicastrin is a crucial member in the membrane-associated γ-secretase complex which is responsible for the specific proteolysis of beta-Amyloid that cause Alzheimer Disease (AD). It is well-established that the stability of the γ-secretase complex is essential for its function (Takasugi et al. Nature. 2003 Mar 27;422(6930):438-41. Epub 2003 Mar 16.) . Coatomer proteins are responsible for the

25   stability and transport of complexes to the membrane. This observation makes AD a disease of complex stability rather than a disease of proteolysis. Databases such as LocusLINK show a linkage between nicastrin and Alzheimer's disease, but not between COPA and Alzheimer's disease. Therefore, this finding represents a new linkage and also a potential new therapeutic modality.

30   Since COPA protein (long form see Figure 24b) clearly stabilizes NCSTN, it is suggested that downregulation of COPA even in combination with other treatment

methods against the gamma-secretase members, will be very efficient in combating AD.

EXAMPLE 12

5   *Enhancing a therapeutic efficacy of IL-24 using a TOSO inhibitor*

The present inventors uncovered that TOSO (GenBank Accession No. NM_005449) is a naturally occurring antisense transcript of IL-24 (GenBank Accession No. NM_006850); overlap alignement between TOSO and IL-24 is provided in Figure 25a. Products of TOSO and IL-24 regulate apoptosis in an

10   opposite manner, as implied from their inverse expression patterns (see Figure 25b). Essentially, while IL-24 induces selective apoptosis in various cancer cells, TOSO is a potent inhibitor of FAS-induced apoptotic signaling through its extracellular domain (see for example information at LocusLINK, which may be obtained by entering the above names to the database).

15   Since TOSO and IL-24 exhibit opposing functions at the protein level, it is logical that expression levels thereof follow an inverse relationship. To ensure such inverse expression patterns, TOSO mRNA functions as a negative regulator (antisense) of IL-24 by preventing or reducing IL-24 mRNA translation.

Due to this sense/antisense agonist/antagonist relationship between TOSO and

20   IL-24 it is highly likely that therapeutic efficacy of IL-24 can be enhanced by inhibiting the expression of TOSO, by using, for example, a small molecule inhibitor to inhibit TOSO expression or expression or activity of endogenous TOSO regulators.

Although the invention has been described in conjunction with specific

25   embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents, patent applications and sequences identified by their accession numbers mentioned in this specification are

30   herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent, patent application or sequence identified by their accession number was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any

reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

179
CD-ROM Content

The following CD-ROMs are attached herewith:

Information provided as: File name/byte size/date of creation /operating system/machine format

CD-ROM1:

1.   AS_patent_data_description.doc/ 2 KB/ July 31, 2003/ Microsoft Windows Internet Explorer/ PC.

2.   seqs_125/ 335,513 Kb/ November 15, 2001/ Microsoft Windows Internet Explorer/ PC.

3.   seqs_133/ 253,406 Kb / April 8, 2003/ Microsoft Windows Internet Explorer/ PC.

CD-ROM2:

1.   alignments_125/ 391,693 Kb/ November 15, 2001/ Microsoft Windows Internet Explorer/ PC.

2.   table_125/ 13,926 Kb/ November 15, 2001/ Microsoft Windows Internet Explorer/ PC.

3.   Table_S1.txt/ 41 Kb bytes/ July 31, 2003/ Microsoft Windows / PC.

4.   Table_S2.txt/ 135 Kb/ July 31,, 2003/ Microsoft Windows / PC.

CD-ROM3:

1.   alignments_133/ 454,180 Kb/ April 8, 2003/ Microsoft Windows Internet Explorer/ PC.

2.   table_133/ 10,741 Kb/ April 8, 2003/ Microsoft Windows Internet Explorer/ PC.

3.   annotations_133 / 5,065 Kb/ May 17, 2004/ Microsoft Windows Internet Explorer/ PC.

4.   pep_seqs_133/6,470 Kb/ May 17, 2004/ Microsoft Windows Internet Explorer/ PC.

5.   Trans2Pep/565 Kb/ May 17, 2004/ Microsoft Windows Internet Explorer/ PC.

WHAT IS CLAIMED IS:

1.   A method of identifying putative naturally occurring antisense transcripts, the method comprising:

(a)   computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b)   identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database according to at least one sequence criterion, thereby identifying putative naturally occurring antisense transcripts, wherein said at least one sequence criterion includes one or more of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

2.   The method of claim 1, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

3.   The method of claim 1, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

4.   The method of claim 1, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

5.   The method of claim 1, wherein said second database is generated by:

(i)   providing a library of expressed polynucleotides;

(ii)   obtaining sequence information of said expressed polynucleotides;

181

(iii)   computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv)   storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

6.   The method of claim 1 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

7.   The method of claim 1 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site, poly(T) head, poly(A) tail, and poly(A) signal.

8.   A kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, said sequence region not being complementary with a naturally occurring antisense transcript.

9.   The kit of claim 8, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

10.   The kit of claim 8, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

11.   The kit of claim 8, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

12.   The kit of claim 8, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

13.   The kit of claim 8, wherein said at least one oligonucleotide is labeled.

14.   The kit of claim 8, wherein said at least one oligonucleotide is attached to a solid substrate.

15.   The kit of claim 14, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

16.   A kit for quantifying at least one mRNA transcript of interest, the kit comprising at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA transcript of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest.

17.   The kit of claim 16, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

18.   The kit of claim 16, wherein said first and second oligonucleotides are single stranded oligonucleotides.

19.   The kit of claim 16, wherein said first and second oligonucleotides are double stranded oligonucleotide.

20.   The kit of claim 16, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

21.   The kit of claim 16, wherein said first and second oligonucleotides are labeled.

183

22.   The kit of claim 16, wherein said first and second oligonucleotides are attached to a solid substrate.

23.   The kit of claim 22, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

24.   A kit for quantifying at least one naturally occurring antisense transcript of interest, the kit comprising at least one oligonucleotide being designed and configured so as to be complementary to a sequence region of the at least one naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript.

25.   The kit of claim 24, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

26.   The kit of claim 24, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

27.   The kit of claim 24, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

28.   The kit of claim 24, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

29.   The kit of claim 24, wherein said at least one oligonucleotide is labeled.

30.   The kit of claim 24, wherein said at least one oligonucleotide is attached to a solid substrate.

31.   The kit of claim 30, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

32.   A method of designing artificial antisense transcripts, the method comprising:

(a)   providing a database of naturally occurring antisense transcripts;

(b)   extracting from said database criteria governing structure and/or function of said naturally occurring antisense transcripts, wherein said criteria governing structure and/or function of said naturally occurring antisense transcripts include one or more of antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.; and

(c)   designing the artificial antisense transcripts according to said criteria.

33.   The method of claim 32, wherein said step of providing said database of naturally occurring antisense transcripts is effected by:

(a)   computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b)   identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database,

185

(c)   storing a sequence of said expressed polynucleotide sequences identified in step (b), thereby providing said database of said naturally occurring antisense transcripts..

34.   The method of claim 33, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

35.   The method of claim 33, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

36.   The method of claim 33, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

37.   The method of claim 33, wherein said second database is generated by:
(i)   providing a library of expressed polynucleotides;
(ii)   obtaining sequence information of said expressed polynucleotides;
(iii)   computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and
(iv)   storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

38.   The method of claim 37, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

186

39.   The method of claim 33, further comprising the step of testing said putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

40.   The method of claim 33 further comprising the step of computationally testing said putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

41.   A computer readable storage medium comprising a database including a plurality of sequences, wherein each sequence is of a naturally occurring antisense transcript.

42.   The computer readable storage medium of claim 41, wherein said database further includes information pertaining to each sequence of said naturally occurring antisense transcripts, said information is selected from the group consisting of related sense gene, antisense length, complementarity length, complementarity position, intron molecules, alternative splicing sites, tissue specificity, pathological abundance, chromosomal mapping, open reading frames, promoters, hairpin structures, helix structures, stem and loops, pseudoknots and tertiary interactions, guanidine and/or cytosine content, guanidine tandems, adenosine content, thermodynamic criteria, RNA duplex melting point, RNA modifications, protein-binding motifs, palindromic sequence and predicted single stranded and double stranded regions.

43.   The computer readable storage medium of claim 41, wherein said database further includes information pertaining to generation of said database and potential uses of said database.

187

44.   A method of generating a database of naturally occurring antisense transcripts, the method comprising:

(a)   computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences;

(b)   identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database so as to identify putative naturally occurring antisense transcripts; and

(c)   storing sequence information of said identified naturally occurring antisense transcripts, thereby generating the database of the naturally occurring antisense transcripts.

45.   The method of claim 44, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

46.   The method of claim 44, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

47.   The method of claim 44, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

48.   The method of claim 44, wherein said second database is generated by:

(i)   providing a library of expressed polynucleotides;

(ii)   obtaining sequence information of said expressed polynucleotides;

(iii)   computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

188

(iv)   storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

49.   The method of claim 48, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

50.   The method of claim 44 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

51.   The method of claim 44 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

52.   A system for generating a database of a plurality of putative naturally occurring antisense transcripts, the system comprising a processing unit, said processing unit executing a software application configured for:

(a)   computationally aligning a first database including sense-oriented polynucleotide sequences with a second database including expressed polynucleotide sequences; and

(b)   identifying expressed polynucleotide sequences from said second database being capable of forming a duplex with at least one sense-oriented polynucleotide sequence of said first database according to at least one sequence criterion, wherein said at least one sequence criterion includes at least one of sequence length, sequence

annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

53.   The system of claim 52, wherein said first database includes sequences of a type selected from the group consisting of genomic sequences, expressed sequence tags, contigs, intron sequences, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

54.   The system of claim 52, wherein said second database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

55.   The system of claim 52, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

56.   The system of claim 52, wherein said second database is generated by:

(i)   providing a library of expressed polynucleotides;

(ii)   obtaining sequence information of said expressed polynucleotides;

(iii)   computationally selecting at least a portion of said expressed polynucleotides according to at least one sequence criterion; and

(iv)   storing said sequence information of said at least a portion of said expressed polynucleotides thereby generating said second database.

57.   The system of claim 56, wherein said at least one sequence criterion for computationally selecting said at least a portion of said expressed polynucleotide is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

58.   The system of claim 52 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form said duplex with said at least one sense oriented polynucleotide sequence under physiological conditions.

59.   The system of claim 52 further comprising the step of computationally testing the putative naturally occurring antisense transcripts according to at least one criterion selected from the group consisting of sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

60.   A method of identifying putative naturally occurring antisense transcripts, the method comprising screening a database of expressed polynucleotides sequences according to at least one sequence criterion, said at least one sequence criterion being selected to identify putative naturally occurring antisense transcripts.

61.   The method of claim 60, wherein said database includes sequences of a type selected from the group consisting of expressed sequence tags, contigs, complementary DNA (cDNA) sequences, pre-messenger RNA (mRNA) sequences and mRNA sequences.

62.   The method of claim 61, wherein an average sequence length of said expressed polynucleotide sequences of said second database is selected from a range of 0.02 to 0.8 Kb.

63.   The method of claim 61, wherein said at least one sequence criterion is selected from the group consisting of sequence length, sequence annotation, sequence information, intron splice consensus site, intron sharing, sequence overlap, rare restriction site , poly(T) head, poly(A) tail, and poly(A) signal.

64.   The method of claim 61 further comprising the step of testing the putative naturally occurring antisense transcripts for an ability to form a duplex with at least one sense oriented polynucleotide sequence under physiological conditions.

65.   A method of quantifying at least one mRNA of interest in a biological sample, the method comprising:

(a)   contacting the biological sample with at least one oligonucleotide capable of binding with the at least one mRNA of interest, wherein said at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the mRNA transcript of interest, said sequence region not being complementary with a naturally occurring antisense transcript; and

(b)   detecting a level of binding between the at least one mRNA of interest and said at least one oligonucleotide to thereby quantify the at least one mRNA of interest in the biological sample.

66.   The method of claim 65, wherein said at least one oligonucleotide is attached to a solid substrate.

67.   The method of claim 66, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

68.   The method of claim 65, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

69.   The method of claim 65, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

70.   The method of claim 65, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

71.   The method of claim 65, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

72.   The method of claim 65, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

73.   A method of quantifying the expression potential of at least one mRNA of interest in a biological sample, the method comprising:

(a)   contacting the biological sample with at least one pair of oligonucleotides including a first oligonucleotide capable of binding the at least one mRNA of interest and a second oligonucleotide being capable of binding a naturally occurring antisense transcript complementary to the mRNA of interest; and

(b)   detecting a level of binding between the at least one mRNA of interest and said first oligonucleotide and a level of binding between said naturally occurring antisense transcript complementary to the mRNA of interest and said second oligonucleotide to thereby quantify the expression potential of the at least one mRNA of interest in the biological sample.

74.   The method of claim 73, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

75.   The method of claim 73, wherein said first and second oligonucleotides are single stranded oligonucleotides.

76.   The method of claim 73, wherein said first and second oligonucleotides are double stranded oligonucleotide.

77.   The method of claim 73, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

78.   The method of claim 73, wherein said first and second oligonucleotides are labeled and whereas step (b) is effected by quantifying said label.

79.   The method of claim 73, wherein said first and second oligonucleotides are attached to a solid substrate.

80.   The method of claim 79, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

81.   A method of quantifying at least one naturally occurring antisense transcript of interest in a biological sample, the method comprising:

(a)   contacting the biological sample with at least one oligonucleotide capable of binding with the at least one naturally occurring antisense transcript of interest, wherein said at least one oligonucleotide is designed and configured so as to be complementary to a sequence region of the naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript; and

(b)   detecting a level of binding between the at least one naturally occurring antisense transcript of interest and said at least one oligonucleotide to thereby quantify the at least one naturally occurring antisense transcript of interest in the biological sample.

82.   The method of claim 81, wherein said at least one oligonucleotide is attached to a solid substrate.

83.   The method of claim 82, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

194

84.   The method of claim 81, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

85.   The method of claim 81, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

86.   The method of claim 81, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

87.   The method of claim 81, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

88.   The method of claim 81, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

89.   A method of identifying a novel drug target, the method comprising:

(a)   determining expression level of at least one naturally occurring antisense transcript of interest in cells characterized by an abnormal phenotype; and

(b)   comparing said expression level of said at least one naturally occurring antisense transcript of interest in said cells characterized by an abnormal phenotype to an expression level of said at least one naturally occurring antisense transcript of interest in cells characterized by a normal phenotype, to thereby identify the novel drug target.

90.   The method of claim 89, wherein said abnormal phenotype of said cells is selected from the group consisting of biochemical phenotype, morphological phenotype and nutritional phenotype.

91.   The method of claim 89, wherein said determining expression level of at least one naturally occurring antisense transcript of interest is effected by at least one oligonucleotide designed and configured so as to be complementary to a sequence

region of said at least one naturally occurring antisense transcript of interest, said sequence region not being complementary with a naturally occurring mRNA transcript.

92.   The method of claim 91, wherein a length of said at least one oligonucleotide is selected from a range of 15-200 nucleotides.

93.   The method of claim 91, wherein said at least one oligonucleotide is a single stranded oligonucleotide.

94.   The method of claim 91, wherein said at least one oligonucleotide is a double stranded oligonucleotide.

95.   The method of claim 91, wherein a guanidine and cytosine content of said at least one oligonucleotide is at least 25 %.

96.   The method of claim 91, wherein said at least one oligonucleotide is labeled and whereas step (b) is effected by quantifying said label.

97.   The method of claim 91, wherein said at least one oligonucleotide is attached to a solid substrate.

98.   The method of claim 97, wherein said solid substrate is configured as a microarray and whereas said at least one oligonucleotide includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

99.   A method of treating or preventing a disease, condition or syndrome associated with an upregulation of a naturally occurring antisense transcript complementary to a naturally occurring mRNA transcript, the method comprising administering a therapeutically effective amount of an agent for regulating expression of the naturally occurring antisense transcript.

196

100.   The method of claim 99, wherein said agent for regulating expression of the naturally occurring antisense transcript is at least one oligonucleotide designed and configured so as to hybridize to a sequence region of said at least one naturally occurring antisense transcript.

101.   The method of claim 100, wherein said at least one oligonucleotide is a ribozyme.

102.   The method of claim 100, wherein said at least one oligonucleotide is a sense transcript.

103.   A method of diagnosing a disease, condition or syndrome associated with a substandard expression ratio of an mRNA of interest over a naturally occurring antisense transcript complementary to the mRNA of interest, the method comprising:

(a)   quantifying expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest;

(b)   calculating the expression ratio of the mRNA of interest over the naturally occurring antisense transcript complementary to the mRNA of interest, thereby diagnosing the disease, condition or syndrome.

104.   The method of claim 103, wherein quantifying said expression level of the mRNA of interest and the naturally occurring antisense transcript complementary to the mRNA of interest is effected by at least one pair of oligonucleotides including a first oligonucleotide capable of binding the mRNA of interest and a second oligonucleotide being capable of binding the naturally occurring antisense transcript complementary to the mRNA of interest.

105.   The method of claim 104, wherein a length of each of said first and second oligonucleotides is selected from a range of 15-200 nucleotides

106.   The method of claim 104, wherein said first and second oligonucleotides are single stranded oligonucleotides.

107.   The method of claim 104, wherein said first and second oligonucleotides are double stranded oligonucleotides.

108.   The method of claim 104, wherein a guanidine and cytosine content of each of said first and second oligonucleotides is at least 25 %.

109.   The method of claim 104, wherein said first and second oligonucleotides are labeled.

110.   The method of claim 104, wherein said first and second oligonucleotides are attached to a solid substrate.

111.   The method of claim 110, wherein said solid substrate is configured as a microarray and whereas each of said first and second oligonucleotides includes a plurality of oligonucleotides each attached to said microarray in a regio-specific manner.

112.   A computer readable storage medium comprising data stored in a retrievable manner, said data including sequence information of co-regulated human polynucleotide sequences as set forth in files seqs_125 and/or seqs_133 of enclosed CD-1, polypeptide sequences encoded by said polynucleotide sequences as set forth in the file pep_seqs_133 of enclosed CD-ROM-3 and sequence annotations as set forth in the file annotations_ 133 of enclosed CD-ROM-3.

113.   A method of modulating an activity or expression of a gene product, the method comprising upregulating or down regulating expression or activity of a naturally occurring antisense transcript of the gene product, thereby modulating the activity or expression of the gene product.

114.   The method of claim 113, wherein said naturally accurring antisense transcript of the gene product is any of the nucleic acid sequences set forth in the file seqs_125 or seqs_133 of the enclosed CD-ROM1.

115.   The method of claim 113, further comprising upregulating or down regulating expression or activity of the gene product.

116.   An isolated polynucleotide comprising any of the nucleic acid sequences set forth in the file seqs_125 or seqs_133 of the enclosed CD-ROM1.

117.   An isolated polypeptide comprising any of the amino acid sequences set forth in the file pep_seqs_133 of enclosed CD-ROM 3.

118.   A method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a naturally occurring antisense transcript of a gene encoding an expression product targeted by the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent, thereby enhancing the therapeutic efficacy of the therapeutic agent.

119.   A method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a naturally occurring antisense transcript of a gene encoding an expression product capable of inhibiting therapeutic activity of the therapeutic agent, prior to, concomitant with, or following provision of the therapeutic agent thereby enhancing the therapeutic efficacy of the therapeutic agent.

120.   A method of enhancing a therapeutic efficacy of a therapeutic agent, the method comprising providing to a subject in need thereof a polynucleotide complementary to a naturally occurring antisense transcript of a gene encoding an expression product capable of upregulating activity or expression of the therapeutic agent prior to, concomitant with, or following provision of the therapeutic agent, thereby enhancing the therapeutic efficacy of the therapeutic agent.

121.   An isolated polynucleotide, comprising a polynucleotide selected according to any of claims 118-120.

122.   The isolated polynucleotide of claim 121, comprising any of the nucleic acid sequences set forth in the file seqs_125 or seqs_133 of the enclosed CD-ROM1, selected according to any of claims 118-120.

Fig. 1

Figure 2

12

32

50

22

18
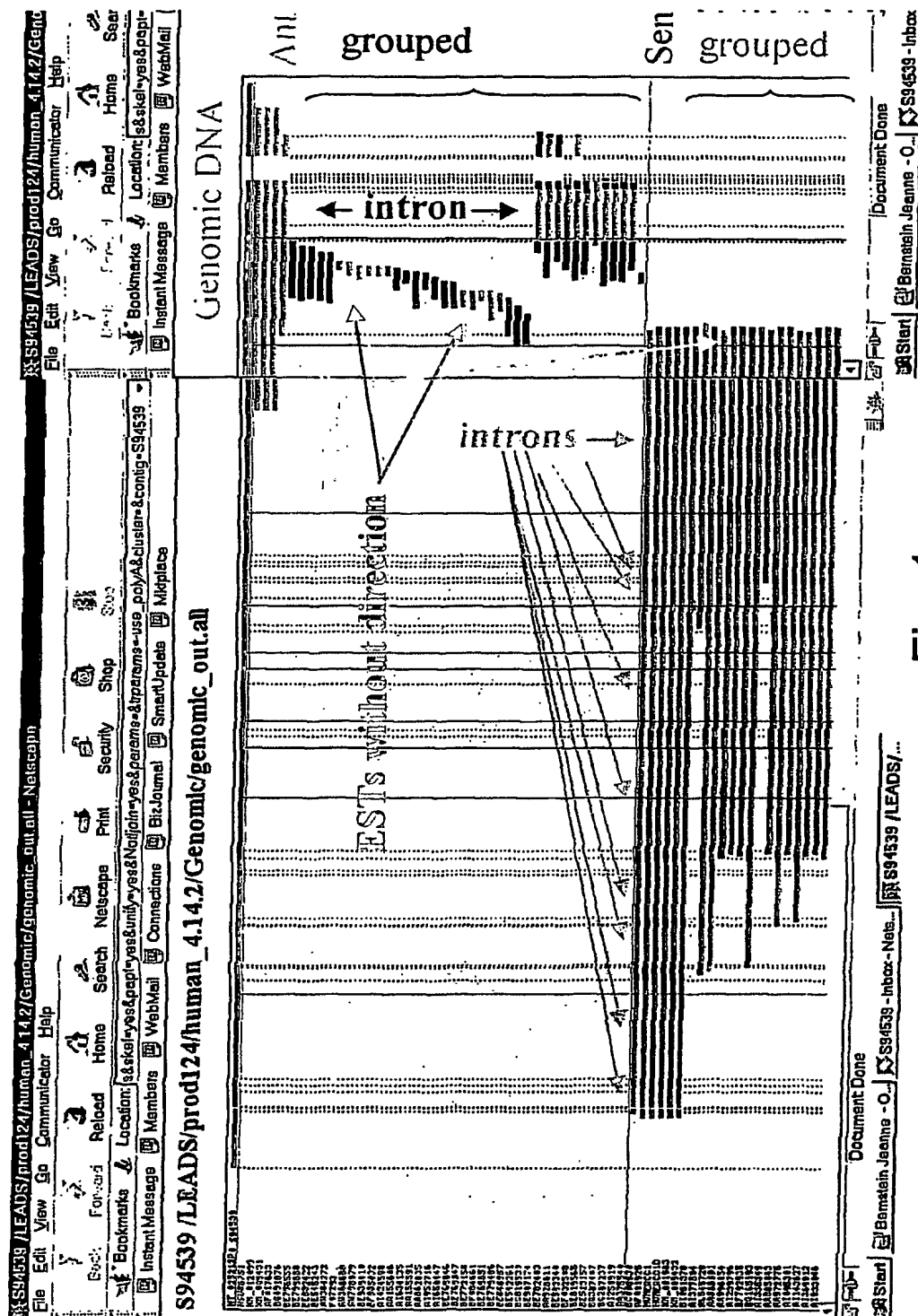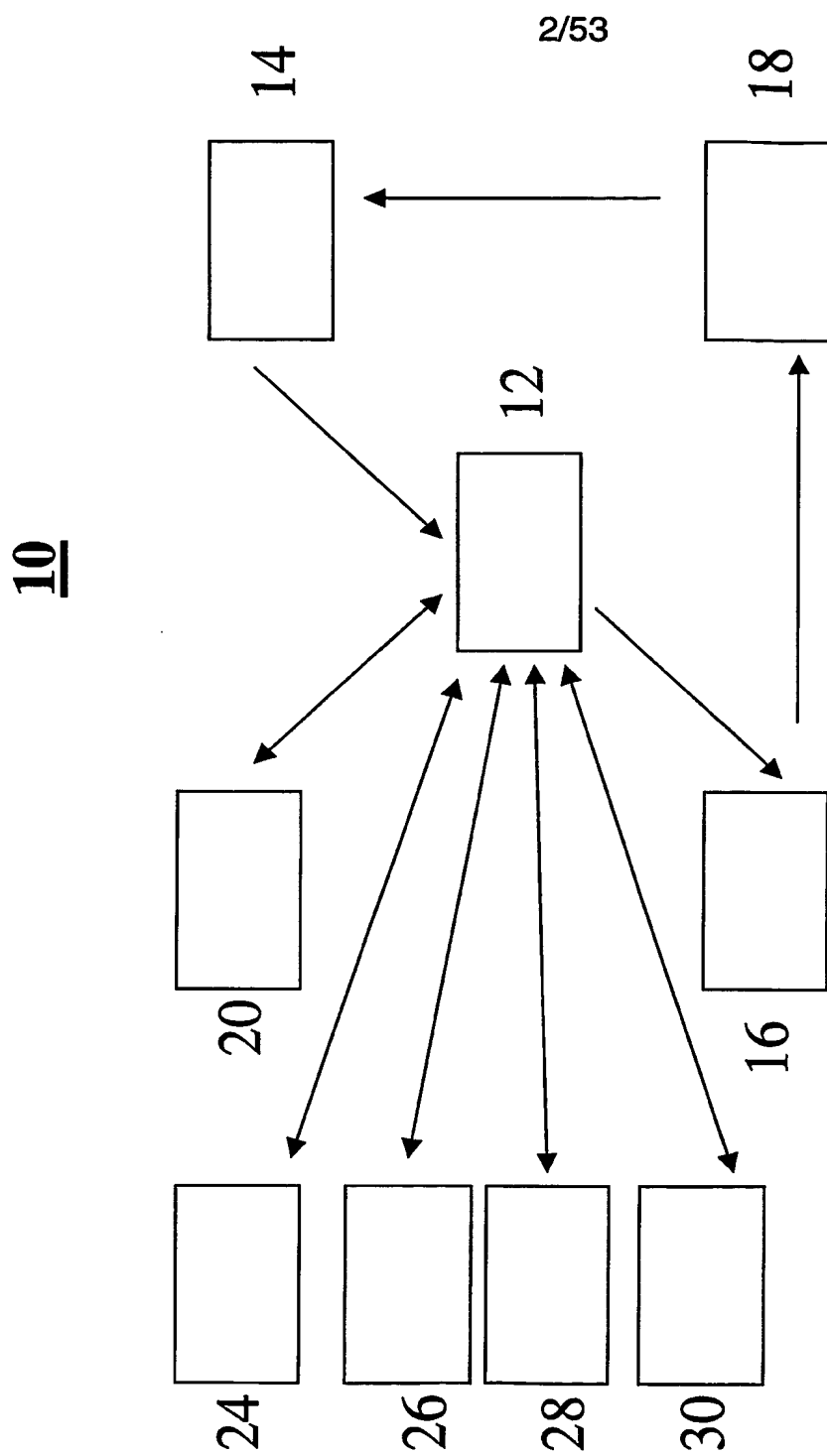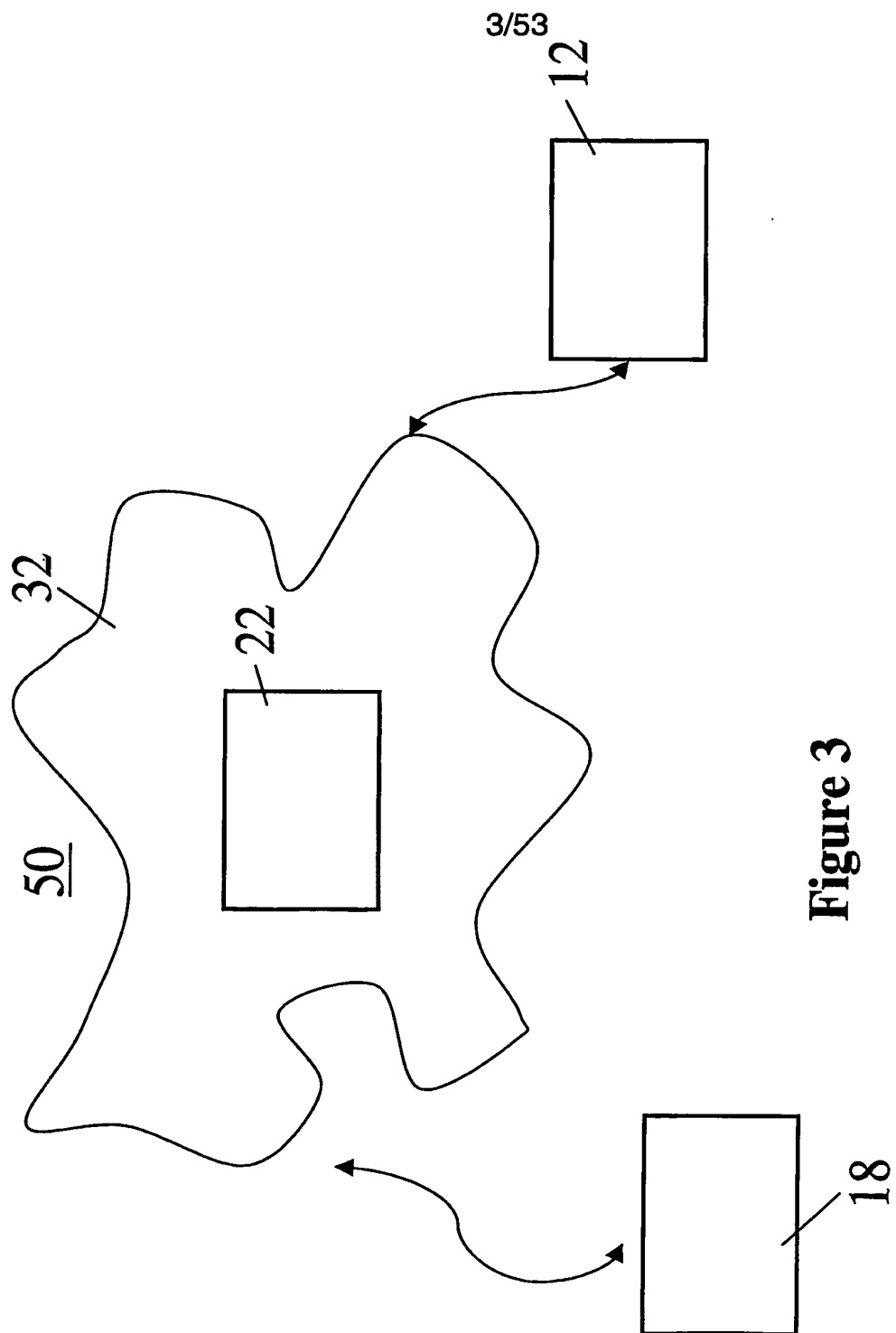
Figure 3

Fig. 4a

```
570_0   AV705532_0       190      Z44352_15       783    OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4b

```
570_1   AV705532_0       190      Z44352_14      1649    OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4c

```
570_2   AV705532_0       190      Z44352_13      1861    OL: 52
Query: 1  ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 52
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 52 ggacccaggatatgagcggaaaacactttctctacttagatacaactttttc 1
```

Fig. 4d

```
571_0   AW070860_0       214      T81142_7       1934    OL: 54
Query: 1    gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1215 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1162
```

```
Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Minus
```

```
Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 760 acttccagagg 750
```

Fig. 4e

```
571_1   AW070860_0       214      T81142_6       2353    OL: 54
Query: 1    gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1215 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1162
```

```
Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Minus
```

```
Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 760 acttccagagg 750
```

```
Score = 22.3 bits (11), Expect = 0.66
Identities = 11/11 (100%)
Strand = Plus / Plus


Query: 100  ggaaaacacac 110
            |||||||||||
Sbjct: 1900 ggaaaacacac 1910
```

Fig. 4f

```
571_2  AW070860_0      214      T81142_4        2500      OL: 54
Query: 1     gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1317 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 1264

Score = 22.3 bits (11), Expect = 0.66
 Identities = 11/11 (100%)
 Strand = Plus / Minus


Query: 146 acttccagagg 156
           |||||||||||
Sbjct: 862 acttccagagg 852


 Score = 22.3 bits (11), Expect = 0.66
 Identities = 11/11 (100%)
 Strand = Plus / Plus


Query: 100  ggaaaacacac 110
            |||||||||||
Sbjct: 2047 ggaaaacacac 2057
```

Fig. 4g

```
571_3  AW070860_0      214      T81142_3        947      OL: 54
Query: 1     gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 224 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 171
```

Fig. 4h

```
571_4  AW070860_0      214      T81142_2        1366      OL: 54
Query: 1     gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 54
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 224 gtaagggaactttggcgacttagtgcgatcactgggagaattgtagagtccact 171


Score = 22.3 bits (11), Expect = 0.66
```

```
Identities = 11/11 (100%)
Strand = Plus / Plus


Query: 100 ggaaaacacac 110
           |||||||||||
Sbjct: 913 ggaaaacacac 923
```

Fig. 4i

```
572_0   BE046369_0      422     W26553_3        1532    OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1481 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1532
```

Fig. 4j

```
572_1   BE046369_0      422     W26553_2        1753    OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1702 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1753
```

Fig. 4k

```
572_2   BE046369_0      422     W26553_1        1832    OL: 52
Query: 1     aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 52
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1781 aatcttcataatccccatgtgtcaaaggagagaccaggtggaggtaactgaa 1832
```

53BP1_76P   53BP1   10394   76P   6837   OL: 3046   OF1: 5463
OF2: 2018

 Score = 1659 bits (837), Expect = 0.0
 Identities = 840/841 (99%)
 Strand = Plus / Minus

Query: 7432 gagacggaatttcgctcttgttgcccaggctggaatgcaatggcacaatctcagctcact
7491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3113 gagacggaatttcgctcttgttgcccaggctggaatgcaatggcacaatctcagctcact
3054

Query: 7492 gcagcgtctgcttcccaggttcaagcaattctcctgtctcagcctcctgagtagctggga
7551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3053 gcagcgtctgcttcccaggttcaagcaattctcctgtctcagcctcctgagtagctggga
2994

Query: 7552 ttacaggcacatgccaccacacctggctaatttttgtatttttagtagaatcgaggtttc
7611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2993 ttacaggcacatgccaccacacctggctaatttttgtatttttagtagaatcgaggtttc
2934

Query: 7612 atcatgttggtcaggctggtctcaaactcctgacttcaggtgatccgcccgcctcggcct
7671 ¦
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2933 atcatgttggtcaggctggtctcaaactcctgacttcaggtgatccgcccgcctcggcct
2874

Query: 7672 cccaaagtgctgggattacaggtgtgagccaccatgcccggcctaagaaatacttttaag
7731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2873 cccaaagtgctgggattacaggtgtgagccaccatgcccggcctaagaaatacttttaag
2814

Query: 7732 tatattttcattagctagaattgcccaatctgtgtaggtataaattacttggtataggga
7791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2813 tatattttcattagctagaattgcccaatctgtgtaggtataaattacttggtataggga
2754

Query: 7792 gagagaaagcctatcttacctgttgctttcttacttggtggtaacatccagcagttagtc
7851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2753 gagagaaagcctatcttacctgttgctttcttacttggtggtaacatccagcagttagtc
2694

Fig. 5a

8/53

```
Query: 7852 tatttataaacataattacttttcacatatgaaccataaaatatttaactttctgctct
7911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2693 tatttataaacataattacttttcacatatgaaccataaaatatttaactttctgctct
2634


Query: 7912 atattgtttgtttaccgctgtatctcccacagcttgaacagtaccaaggtacgtagtagg
7971
             ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2633 atattgtttgtctaccgctgtatctcccacagcttgaacagtaccaaggtacgtagtagg
2574


Query: 7972 tgctcaataaatgactattgaataaatgaacatatccaacaaatgttctcaatgtaaagg
8031
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2573 tgctcaataaatgactattgaataaatgaacatatccaacaaatgttctcaatgtaaagg
2514


Query: 8032 atcagagatgccacatgttctccttgatgggagagacccttccacatgggaatgatggga
8091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2513 atcagagatgccacatgttctccttgatgggagagacccttccacatgggaatgatggga
2454


Query: 8092 aggagttgtactcctggatgttcagtaactgcttctaggagaaaaggtagagtcctatca
8151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2453 aggagttgtactcctggatgttcagtaactgcttctaggagaaaaggtagagtcctatca
2394


Query: 8152 ctaagccgcagatatttatttgtgtgtggctagaatgggatgttttgaatcttctgttac
8211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2393 ctaagccgcagatatttatttgtgtgtggctagaatgggatgttttgaatcttctgttac
2334


Query: 8212 aaccttgggaacgtggctgttatttcaatttatgagccagaaattttcacatcccgaaac
8271
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2333 aaccttgggaacgtggctgttatttcaatttatgagccagaaattttcacatcccgaaac
2274


Query: 8272 t 8272
             |
Sbjct: 2273 t 2273
```

Score = 1655 bits (835), Expect = 0.0
Identities = 849/856 (99%)
Strand = Plus / Minus

Fig. 5a continued

```
Query:  5903 agatattgctttaggggtatttgatgtggtggtgacggacccctcatgcccagcctcggt
        5962
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4642 agatattgctttaggggtatttgatgtggtggtgacggacccctcatgcccagcctcggt
        4583


Query:  5963 gctgaagtgtgctgaagcattgcagctgcctgtggtgtcacaagagtgggtgatccagtg
        6022
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4582 gctgaagtgtgctgaagcattgcagctgcctgtggtgtcacaagagtgggtgatccagtg
        4523


Query:  6023 cctcattgttggggagagaattggattcaagcagcatccaaaatataaacacgattatgt
        6082
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4522 cctcattgttggggagagaattggattcaagcagcatccaaaatataaacacgattatgt
        4463


Query:  6083 ttctcactaaagatacttggtcttactggtttttattccctgctatcgtggagattgtgtt
        6142
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4462 ttctcactaaagatacttggtcttactggtttttattccctgctatcgtggagattgtgtt
        4403


Query:  6143 ttaaccaggttttaaatgtgtcttgtgtgtaactggattccttgcatggatcttgtatat
        6202
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4402 ttaaccaggttttaaatgtgtcttgtgtgtaactggattccttgcatggatcttgtatat
        4343


Query:  6203 agttttatttgctgaacttttatgataaaataaatgttgaatctctttggttgtagtaac
        6262
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4342 agttttatttgctgaacttttatgataaaataaatgttgaatctctttggttgtagtaac
        4283


Query:  6263 tgggatttcttcatctgnnnnnnnngagcttaatctcagaacaaatgacaagacatagtac
        6322
             |||||||||||||||||         |||||||||||||||||||||||||||||||||||
Sbjct:  4282 tgggatttcttcatctgttttttgagcttaatctcagaacaaatgacaagacatagtac
        4223


Query:  6323 tttctctgagtctttcaacaggcttattcacttacggaggacagctcaccaaggaaattg
        6382
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4222 tttctctgagtctttcaacaggcttattcacttacggaggacagctcaccaaggaaattg
        4163


Query:  6383 aaaagttaagagtgaactttattctgtggcatcattcccaaaaggttattccagggtgtc
        6442
```

Fig. 5a continued

```
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4162  aaaagttaagagtgaactttattctgtggcatcattcccaaaaggttattccagggtgtc
4103


Query:  6443  taaaatgctatgcttgcagaaactcagtttaaggtaggtgaaggcccagattaacagttg
6502
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4102  taaaatgctatgcttgcagaaactcagtttaaggtaggtgaaggcccagattaacagttg
4043


Query:  6503  tgccaaaagttgagtggaattgggcacagctctgtttcctgacagttaaaaaagacctca
6562
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4042  tgccaaaagttgagtggaattgggcacagctctgtttcctgacagttaaaaaagacctca
3983


Query:  6563  tgctctctctctgagctgagatcacagctcacctgtgggtactccccaactcttagagct
6622
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3982  tgctctctctctgagctgagatcacagctcacctgtgggtactccccaactcttagagct
3923


Query:  6623  aaagggagaacgaaaggaccaactgccatgaagggacagtgaccataagcttgatggaat
6682
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3922  aaagggagaacgaaaggaccaactgccatgaagggacagtgaccataagcttgatggaat
3863                  ;


Query:  6683  gaccttccgtaagataaacatgggaagcacaagtgagaacacctggaaatgttacacgtt
6742
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3862  gaccttccgtaagataaacatgggaagcacaagtgagaacacctggaaatgttacacgtt
3803


Query:  6743  ctagtcaaagacccaa  6758
                     ||||||||||||||||
Sbjct:  3802  ctagtcaaagacccaa  3787


 Score = 1211 bits (611), Expect = 0.0
 Identities = 625/632 (98%)
 Strand = Plus / Minus


Query:  6778  gtcacaatagctggaagcagttccttcccttcctctggcatcactgatccctgcatggct
6837
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3767  gtcacaatagctggaagcagttccttcccttcctctggcatcactgatccctgcatggct
3708
```

Fig. 5a continued

Query: 6838 tctcattctctaaagcagggggtcaacaaggnnnnnnnnnctgtaaagggtcaaagagtaaat
6897
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||
Sbjct: 3707 tctcattctctaaagcagggggtcaacaaggttttttttctgtaaagggtcaaagagtaaat
3648

Query: 6898 atttcaggctttgtgggccatttgatccatcacaactactcgcctttgctgtgagggcat
6957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3647 atttcaggctttgtgggccatttgatccatcacaactactcgcctttgctgtgagggcat
3588

Query: 6958 gaaagcaaccatagacaatgagtaaacaaatgggcacggctgtgtttcagtaaaactgta
7017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3587 gaaagcaaccatagacaatgagtaaacaaatgggcacggctgtgtttcagtaaaactgta
3528

Query: 7018 caaaaacagacagcaggccatagtttgccagctcctgctccagagacagcagtggaaagg
7077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3527 caaaaacagacagcaggccatagtttgccagctcctgctccagagacagcagtggaaagg
3468

Query: 7078 gtgatctttagttgataatagcagggaataagttgtcagagcttcccagtgtgtgtagaa
7137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3467 gtgatctttagttgataatagcagggaataagttgtcagagcttcccagtgtgtgtagaa
3408

Query: 7138 tatgtagtgatgaaaaccagatgcagtgactataacctgatgccagaacactgcattctt
7197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3407 tatgtagtgatgaaaaccagatgcagtgactataacctgatgccagaacactgcattctt
3348

Query: 7198 tttcagtttggagggcgttgttcagtgaatatttctttttacttacactgatatgaatat
7257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3347 tttcagtttggagggcgttgttcagtgaatatttctttttacttacactgatatgaatat
3288

Query: 7258 tgattaccagtgatggctgggccatattaagataacttcaaccccctatggtttgtgtaag
7317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3287 tgattaccagtgatggctgggccatattaagataacttcaaccccctatggtttgtgtaag
3228

Query: 7318 atgggtaattgggcctgcaatcttcagtatttaaaaatctaacaacttgatctcaatttt
7377

Fig. 5a continued

```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3227 atgggtaattgggcctgcaatcttcagtatttaaaaatctaacaacttgatctcaatttt
3168


Query: 7378 ttcttaaggacctttttcttggagaataatac 7409
           ||||||||||||||||||||||||||||||||
Sbjct: 3167 ttcttaaggacctttttcttggagaataatac 3136
```

  Score = 404 bits (204), Expect = e-115
  Identities = 204/204 (100%)
  Strand = Plus / Minus


```
Query: 5556 cagtgtaacacagcttaccagtgtcttctaattgcggatcagcattgtcgaacccggaag
5615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6169 cagtgtaacacagcttaccagtgtcttctaattgcggatcagcattgtcgaacccggaag
6110


Query: 5616 tacttcctgtgccttgccagtgggattccttgtgtgtctcatgtctgggtccatgatagt
5675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6109 tacttcctgtgccttgccagtgggattccttgtgtgtctcatgtctgggtccatgatagt
6050


Query: 5676 tgccatgccaaccagctccagaactaccgtaattatctgttgccagctgggtacagcctt
5735
           '|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6049 tgccatgccaaccagctccagaactaccgtaattatctgttgccagctgggtacagcctt
5990


Query: 5736 gaggagcaaagaattctggactgg 5759
           ||||||||||||||||||||||||
Sbjct: 5989 gaggagcaaagaattctggactgg 5966
```

  Score = 291 bits (147), Expect = 1e-80
  Identities = 147/147 (100%)
  Strand = Plus / Minus


```
Query: 5758 ggcaaccccgtgaaaatcctttccagaatctgaaggtactcttggtatcagaccaacagc
5817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5159 ggcaaccccgtgaaaatcctttccagaatctgaaggtactcttggtatcagaccaacagc
5100


Query: 5818 agaacttcctggagctctggtctgagatcctcatgactggtggtgcagcctctgtgaagc
5877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5099 agaacttcctggagctctggtctgagatcctcatgactggtggtgcagcctctgtgaagc
5040
```

Fig. 5a continued

Query: 5878 agcaccattcaagtgcccataacaaag 5904
               |||||||||||||||||||||||||||||
Sbjct: 5039 agcaccattcaagtgcccataacaaag 5013

  Score = 281 bits (142), Expect = 9e-78
  Identities = 142/142 (100%)
  Strand = Plus / Minus

Query: 8920 ctgcccagagttccaccagcctgggtatagtatttgttataatctagtcgtaacagtagt
8979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2274 ctgcccagagttccaccagcctgggtatagtatttgttataatctagtcgtaacagtagt
2215

Query: 8980 tgagccaaatctgagttgatctgatgattccgaacactggagagaatcttgaacaggagt
9039
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2214 tgagccaaatctgagttgatctgatgattccgaacactggagagaatcttgaacaggagt
2155

Query: 9040 gaagactggcggctaaagccct 9061
              ||||||||||||||||||||||
Sbjct: 2154 gaagactggcggctaaagccct 2133

  Score = 226 bits (114), Expect = 5e-61
  Identities = 117/118 (99%)
  Strand = Plus / Minus

Query: 9673 ccttcacgagaatgctcagctgggcggctccacgctcatccagtgggcctaggttctgac
9732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2135 ccttcacgagaatgctcagctgggcggctccacgctcatccagtgggcctaggttctgac
2076

Query: 9733 tgaccagcgaacaaaaactgtgacagagatctaggatttcattcaggcagtgaaacac 9790
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2075 tgaccagcaaacaaaaactgtgacagagatctaggatttcattcaggcagtgaaacac 2018

  Score = 190 bits (96), Expect = 3e-50
  Identities = 96/96 (100%)
  Strand = Plus / Minus

Query: 5463 gaatttttggaaattcctcctttcaacaagcagtatacagaatcccagcttcgagcagga
5522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6812 gaatttttggaaattcctcctttcaacaagcagtatacagaatcccagcttcgagcagga
6753

Fig. 5a continued

14/53

```
Query: 5523 gctggctatatccttgaagatttcaatgaagcccag 5558
             ||||||||||||||||||||||||||||||||||||
Sbjct: 6752 gctggctatatccttgaagatttcaatgaagcccag 6717


  Score = 52.0 bits (26), Expect = 2e-08
  Identities = 26/26 (100%)
  Strand = Plus / Minus


Query: 7668 gcctcccaaagtgctgggattacagg 7693
             ||||||||||||||||||||||||||
Sbjct: 5312 gcctcccaaagtgctgggattacagg 5287
```

Fig. 5a continued

CIDEB1_BLTR2   CIDEB1   2289 BLTR2 6530 OL: 2254 OF1: 17   OF2:
         1

Score = 2727 (753.8 bits), Expect = 0.0, Sum P(13) = 0.0
Identities = 547/549 (99%), Positives = 547/549 (99%),   Strands Minus /
Plus

```
Query:  2250 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT
        2191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT 60

Query:  2190 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
        2131
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
        120

Query:  2130 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
        2071
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
        180

Query:  2070 TGGGAGGCTAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
        2011
             .LLLLLLLL .LLLLLLLLLJJJJJJJJJ|||||LLLLLLLLLLLLLLLLLLLLJJJJJJJJJ.
Sbjct:   181 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
        240

Query:  2010 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
        1951
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   241 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
        300  !

Query:  1950 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
        1891
             ·||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   301 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
        360

Query:  1890 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
        1831
             .||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
        420

Query:  1830 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
        1771
             |||||||||||||||||||||||||||||||||||||||||||||·||||||||||||||||
Sbjct:   421 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
        480

Query:  1770 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC
        1711
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||JJJJJJJ
Sbjct:   481 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC
        540
```

Fig. 5b

Query: 1710 TCCCTGAGT 1702
               ||||||| |
Sbjct: 541 TCCCTGAAT 549

 Score = 1322 (365.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 266/268 (99%), Positives = 266/268 (99%),   Strands Minus /
Plus

Query: 757 CCTGTAGGCCCAGAAGGATGTCGGTCTGCTACCGTCCCCCAGGGAACGAGACACTGCTGA
698
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5426 CCTGTAGGCCCAGAAGGATGTCGGTCTGCTACCGTCCCCCAGGGAACGAGACACTGCTGA
5485

Query: 697 GCTGGAAGACTTCGCGGGCCACAGGCACAGCCTTCCTGCTGCTGGCGGCGCTGCTGGGGC
638
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5486 GCTGGAAGACTTCGCGGGCCACAGGCACAGCCTTCCTGCTGCTGGCGGCGCTGCTGGGGC
5545

Query: 637 TGCCTGGCAACGGCTTCGTGGTGTGGAGCTTGGCGGGCTGGCAGCCTGCACGGGGGCGAC
578
               |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct: 5546 TGCCTGGCAACGGCTTCGTGGTGTGGAGCTTGGCGGGCTGGCGGCCTGCACGGGGGCGAC
5605

Query: 577 CGCTGGCGGCCACGCTTGTGCTGCACCTGGCGCTGGCCGACGGCGCGGTGCTGCTGCTCA
518
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5606: CGCTGGCGGCCACGCTTGTGCTGCACCTGGCGCTGGCCGACGGCGCGGTGCTGCTGCTCA
5665

Query: 517 CGCCGCTCTTTGTGGCCTTCCTGACCGG 490
               ||||||||||||||||||||||||||| |
Sbjct: 5666 CGCCGCTCTTTGTGGCCTTCCTGACCCG 5693

 Score = 1316 (363.8 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 264/265 (99%), Positives = 264/265 (99%),   Strands Minus /
Plus

Query: 421 CAAGCGTGCTGCTCACCGGCCTGCTCAGCCTGCAGCGCTGCCTCGCAGTCACCCGCCCCT
362
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5762 CCAGCGTGCTGCTCACCGGCCTGCTCAGCCTGCAGCGCTGCCTCGCAGTCACCCGCCCCT
5821

Query: 361 TCCTGGCGCCTCGGCTGCGCAGCCCGGCCCTGGCCCGCCGCCTGCTGCTGGCGGTCTGGC
302
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5822 TCCTGGCGCCTCGGCTGCGCAGCCCGGCCCTGGCCCGCCGCCTGCTGCTGGCGGTCTGGC
5881

Query: 301 TGGCCGCCCTGTTGCTCGCCGTCCCGGCCGCCGTCTACCGCCACCTGTGGAGGGACCGCG
242
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5882 TGGCCGCCCTGTTGCTCGCCGTCCCGGCCGCCGTCTACCGCCACCTGTGGAGGGACCGCG
5941

Fig. 5b continued

```
Query:   241  TATGCCAGCTGTGCCACCCGTCGCCGGTCCACGCCGCCGCCCACCTGAGCCTGGAGACTC
182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  5942  TATGCCAGCTGTGCCACCCGTCGCCGGTCCACGCCGCCGCCCACCTGAGCCTGGAGACTC
6001

Query:   181  TGACCGCTTTCGTGCTTCCTTTCGG  157
              |||||||||||||||||||||||||
Sbjct:  6002  TGACCGCTTTCGTGCTTCCTTTCGG  6026
```

 Score = 920 (254.3 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 188/193 (97%), Positives = 188/193 (97%),   Strands Minus /
Plus

```
Query:  1708  CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
1649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   724  CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783

Query:  1648  GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   784  GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
843

Query:  1588  AAAGGTGAATCGGCCGATGTCCTTGCTGTGCTTGGGGCTCTCCCGTCCAGGCCCATATGA
1529
              ||||||||||||| ||||||||||||||||||||| |||||||||| | ||||||||
Sbjct:   844  AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCCATATGA
903

Query:  1528  CAGCACTCCACTC  1516
              |||||||||||||
Sbjct:   904  CAGCACTCCACTC  916
```

 Score = 753 (208.2 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 157/165 (95%), Positives = 157/165 (95%),   Strands Minus /
Plus

```
Query:  1529  ACAGCACTCCACTCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCATCAGGCACG
1470
              | ||  |||     |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1139  ATAGGCCTCTTACCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCATCAGGCACG
1198

Query:  1469  TGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATCCTCCTCTA
1410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1199  TGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATCCTCCTCTA
1258

Query:  1409  GCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCTT  1365
              |||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  1259  GCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCCT  1303
```

 Score = 746 (206.2 bits), Expect = 0.0, Sum P(13) = 0.0

Fig. 5b continued

Identities = 150/151 (99%), Positives = 150/151 (99%),   Strands Minus /
Plus

Query: 1369 TGCTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
1310
             I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 2146 TACTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
2205

Query: 1309 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
1250
             IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 2206 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
2265

Query: 1249 AAACTCCGAGCTTATATTAGATACTGACCTG 1219
             IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 2266 AAACTCCGAGCTTATATTAGATACTGACCTG 2296

 Score = 737 (203.7 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 148/150 (98%), Positives = 149/150 (99%),   Strands Minus /
Plus

Query: 1118 CTGCTCTTTCCTTCCTCCTTGGTCGGAGGAGGGGCTGGCTCACTGCTCTGGCTTCATTTT
1059
             IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
·Sbjct: 3257: CTGCTCTTTCCTTCCTCCTTGGTCGGAGGAGGGGCTGGCTCACTGCTCTGGCTTCATTTT
3316

Query: 1058 CCAGAGCTGCCTGCTGCAGTCACACTTAGGTCATCTTCTCTCACTTTTCTCCTTTTGCCG
999
             IIIIIII·IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 3317 CCAGAGCTGCCTGCTGCAGTCACACTTAGGTCATCTTCTCTCACTTTTCTCCTTTTGCCG
3376

Query: 998 ATTAGTGGACGTGACAGAGATGTGAATRTG 969
             IIIIIIIIIIIIIIIIIIIIIIIIII+ I
Sbjct: 3377 ATTAGTGGACGTGACAGAGATGTGAATGGG 3406

 Score = 714 (197.4 bits), Expect = 0.0, Sum P(13) = 0.0
 Identities = 146/150 (97%), Positives = 146/150 (97%),   Strands Minus /
Plus

Query: 166 TTCCTTTCGGCTGATGCTCGGCTGCTACAGCGTGACGCTGGCACGGCTGCGGGGCGCCCG
107
             I I II   IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 6018 TCCTTTCGGGCTGATGCTCGGCTGCTACAGCGTGACGCTGGCACGGCTGCGGGGCGCCCG
6077

Query: 106 CTGGGGCTCCGGGCGGCACGGGGCGCGGGTGGGCCGGCTGGTGAGCGCCATCGTGCTTGC 47
             IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 6078 CTGGGGCTCCGGGCGGCACGGGGCGCGGGTGGGCCGGCTGGTGAGCGCCATCGTGCTTGC
6137

Query: 46 CTTCGGCTTGCTCTGGGCCCCCTACCACGC 17
             IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct: 6138 CTTCGGCTTGCTCTGGGCCCCCTACCACGC 6167

Fig. 5b continued

Score = 638 (176.4 bits), Expect = 0.0, Sum P(13) = 0.0
Identities = 130/133 (97%), Positives = 130/133 (97%), Strands Minus /
Plus

Query: 962 GGGCAGGGGATGTCCTTTGATGGCATCAAGACTTTAGCTTCTGGTGCGCTGTGTCCCAGC
903
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3404 GGGGCAGGGGATGTCCTTTGATGGCATCAAGACTTTAGCTTCTGGTGCGCTGTGTCCCAGC
3463

Query: 902 TCTGATTTCAGTTGCAGCCGTGATGGACAGTTGCATGGAAGCTGAGACTCTCACTGACAG
843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3464 TCTGATTTCAGTTGCAGCCGTGATGGACAGTTGCATGGAAGCTGAGACTCTCACTGACAG
3523

Query: 842 TGAAACCCTCAAA 830
              |||||||||||||
Sbjct: 3524 TGAAACCCTCAAA 3536

Score = 537 (148.4 bits), Expect = 0.0, Sum P(13) = 0.0
Identities = 109/111 (98%), Positives = 109/111 (98%), Strands Minus /
Plus

Query: 1227 ACTGACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
1168
              | | |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2590 AGTCACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
2649

Query: 1167 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCT 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2650 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCT 2700

Score = 394 (108.9 bits), Expect = 0.0, Sum P(13) = 0.0
Identities = 82/86 (95%), Positives = 82/86 (95%), Strands Minus / Plus

Query: 839 AACCCTCAAAATGAACACAATCCCTGCTTTCCTGCCAAGGATCCTTGTAGGGTCCCCCAG
780
              || || |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3526 AAACCCTCAAATGAACACAATCCCTGCTTTCCTGCCAAGGATCCTTGTAGGGTCCCCCAG
3585

Query: 779 CTTCCCCACTTTTTTTCTGTGTCCTG 754
              ||||||||||||||||||||||||||
Sbjct: 3586 CTTCCCCACTTTTTTTCTGTGTCCTG 3611

Score = 370 (102.3 bits), Expect = 0.0, Sum P(13) = 0.0
Identities = 74/74 (100%), Positives = 74/74 (100%), Strands Minus / Plus

Query: 493 CCGGCAGGCCTGGCCGCTGGGCCAGGCGGGCTGCAAGGCGGTGTACTACGTGTGCGCGCT
434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 5691 CCGGCAGGCCTGGCCGCTGGGCCAGGCGGGCTGCAAGGCGGTGTACTACGTGTGCGCGCT
5750

Query: 433 CAGCATGTACGCCA 420
              ||||||||||||||
Sbjct: 5751 CAGCATGTACGCCA 5764

Fig. 5b continued

```
CIDEB2_BLTR2      CIDEB2      1511 BLTR2 6530 OL:    1410 OF1: 1      OF2:
        1

 Score = 2736 (756.0 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 548/549 (99%), Positives = 548/549 (99%),   Strands Minus /
 Plus

Query: 1384 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT
 1325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    1 TTTTGTTAGTTTGAGGGGAAGGGTATGAAGACAGATCTCAAGGTAAAGTCAGAGAGGGCT 60

Query: 1324 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
 1265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   61 GTCATCAGTATGCTGGGGAGTTTAGGGACAGGAGGCATTGGTAGGGGATTAGATGTAGCA
 120

Query: 1264 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
 1205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  121 GCAGTCAGGCTGGGATCAAGATGCCTGGGGGACATCTTGATCTTGGCCTTTCAGGGCAAG
 180

Query: 1204 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
 1145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  181 TGGGAGGCCAGAAAGGTGGCTAGGAAAGAACAGCATTCTTCAGGTAAGGGTATAGACTTG
 240

Query: 1144 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
 1085
            : ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  241 GGATGTGAGGCGTTATGCTGAAAGGTTCTGTCACGAGGGGATCAGAGGACAGTGGGGAAA
 300

Query: 1084 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
 1025
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  301 TTGGGTGGGTTATCTAGCCTGTACTGTCTGCAGGTCCTGAAATTTGATGCTGTCATAGTC
 360

Query: 1024 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
  965
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  361 TTTGCAGTGGGTCGGTTGGAATGATTCTGGGGGCAGAAGCTCAGAGCCCCTTAGTAGGAA
 420

Query:  964 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
  905
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  421 TGGAGGCGGCCCTTCTGCTGCCACTGCTCAGCCCCCTCCACTGCATGACGAAGGGTGGAG
 480

Query:  904 GAAATTCCCAGCAACATATGGCCCAGGCCTTGCAGCAGTGTGGAGGTCCAACGAAGGAGC
  845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  481 GAAATTCCCAGCAACATATGGCCCA(                              :GAAGGAGC
 540
```

Fig. 5c

```
Query:    844 TCCCTGAGT 836
              IIIIIII I
Sbjct:    541 TCCCTGAAT 549


 Score = 1787 (493.8 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 359/361 (99%), Positives = 359/361 (99%),   Strands Minus /
Plus

Query:    361 ACTGACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
302
              I I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2590 AGTCACCTGAGTAAGTCACTGGGGTTCAGAGCTGAGAGGTACTCCATGGTGGACCGGAGA
2649

Query:    301 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCTTTAGTGGTG
242
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2650 GTTCCTTCCCTGGAACTTCTGGGCTGGGTGGTTCTCTCCTGTGCTGGGGCTTTAGTGGTG
2709

Query:    241 TTTTCTGTTACAAACCTGGGATCTCAGCCCAGGACAAGGTGGGAATGAGTCAAGCCTGGA
182
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2710 TTTTCTGTTACAAACCTGGGATCTCAGCCCAGGACAAGGTGGGAATGAGTCAAGCCTGGA
2769

Query:    181 CTCTGGCCCCCCTGCCTGGCCAGTAAGAAGGGCAAAGTCCAAGGGGAGGGATGAGGGAGG
122
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2770 CTCTGGCCCCCCTGCCTGGCCAGTAAGAAGGGCAAAGTCCAAGGGGAGGGATGAGGGAGG
2829

Query:    121 GGCCAGATGGGGTCCTGGAGGAAGAATTGCCTGGCAAAAGCCATTGGAGCTTGTATGTGT 62
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2830 GGCCAGATGGGGTCCTGGAGGAAGAATTGCCTGGCAAAAGCCATTGGAGCTTGTATGTGT
2889

Query:     61 GTCTTTGGTGATGACATGTGTTGTGAGGGTAGATGGGAACCATGTAAAAGGATGAAATGT 2
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:   2890 GTCTTTGGTGATGACATGTGTTGTGAGGGTAGATGGGAACCATGTAAAAGGATGAAATGT
2949

Query:      1 G 1
              I
Sbjct:   2950 G 2950


 Score = 965 (266.6 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 193/193 (100%), Positives = 193/193 (100%),   Strands Minus /
Plus

Query:    842 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Sbjct:    724 CCTGAGTACTTTCTTTGGGCCAAGTCCTTGAAAGTCACAACTCATAGAGTAGAGCCCGTA
783

Query:    782 GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
723
```

Fig. 5c continued

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  784 GAATGTGGCTTTGACATTCAGGCTGCCAAAGAGGTCTCGAGGGTTTTGCTTGTACACGTC
843

Query:  722 AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCATATGA
663
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  844 AAAGGTGAATCGGGCGATGTCCTTGCTGTGCTTGGGCCTCTCCCGTCCCAGGCCATATGA
903

Query:  662 CAGCACTCCACTC 650
            |||||||||||||
Sbjct:  904 CAGCACTCCACTC 916
```

 Score = 757 (209.2 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 161/173 (93%), Positives = 161/173 (93%),   Strands Minus /
Plus

```
Query:  671 GCCATATGACAGCACTCCACTCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCAT
612
            |||  |  | ||  |||    ||||||||||||||||||||||||||||||||||||||||
Sbjct: 1131 GCCCCAGTATAGGCCTCTTACCCTTGTAGGGCTCCAGCTCTGACCAGACTGCAACACCAT
1190

Query:  611 CAGGCACGTGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATC
552
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1191 CAGGCACGTGTCATCCTCCAGCAGCTGGAAGAAGTCCTCACTGTCCACTGCAGTTCCATC
1250

Query:  551 CTCCTCTAGCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCTT 499
            ||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct: 1251 CTCCTCTAGCACCAGGGTTAGCACTCCATTCAGCAGTAGGGTCTCCAATGCCT 1303
```

 Score = 746 (206.1 bits), Expect = 0.0, Sum P(5) = 0.0
 Identities = 150/151 (99%), Positives = 150/151 (99%),   Strands Minus /
Plus

```
Query:  503 TGCTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
444
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2146 TACTTTGGCTAGCAGCTCCTGGCGGGTGGCAGCTGTCAGGCCTTTCCGGATGGTCCGCTT
2205

Query:  443 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2206 GTGATCACAGACACGGAAAGGTCGCTGGGGTGGTGGAGCTGAGGTCCAGACCCTCCGTCC
2265

Query:  383 AAACTCCGAGCTTATATTAGATACTGACCTG 353
            |||||||||||||||||||||||||||||||
Sbjct: 2266 AAACTCCGAGCTTATATTAGATACTGACCTG 2296
```

Fig. 5c continued

APAF1_EB1   APAF1   7042   EB1a   1752   OL: 141   OF1: 6889
OF2: 1612

 Score = 705 (194.8 bits), Expect = 3.9e-52, P = 3.9e-52
 Identities = 141/141 (100%), Positives = 141/141 (100%),   Strands Minus /
Plus

Query: 7029 TGTTTTTCAAAACAATTTTGTGAATTTTATTTTTACAAAAATTTTTTAAATTCATATTTT
6970

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1612 TGTTTTTCAAAACAATTTTGTGAATTTTATTTTTACAAAAATTTTTTAAATTCATATTTT
1671

Query: 6969 AAAATGTATACCAAGGCAAAAAAATCATATAAGCTATATCATAAATACAAGAGTTTCAAA
6910

             |||||||||||||||||||||||||||||||||.|||||||||||||||||||||||||||
Sbjct: 1672 AAAATGTATACCAAGGCAAAAAAATCATATAAGCTATATCATAAATACAAGAGTTTCAAA
1731

Query: 6909 ACATACAAGAGACATATAATG 6889
             ||||||||||||||||||||||
Sbjct: 1732 ACATACAAGAGACATATAATG 1752

Fig. 5d

```
Hum_AChR_MINK2   AChR      2457    MINK2        4863      OL: 236  OF1: 2175
OF2: 4583
```

```
 Score =  218 bits (110), Expect = 2e-59
 Identities = 110/110 (100%)
 Strand = Plus / Minus
```

```
Query: 2254 aagggttacttgctgctcacactatatacagatgcaagcaaggggcgtggagagtgaggg
2313
            .   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4787 aagggttacttgctgctcacactatatacagatgcaagcaaggggcgtggagagtgaggg
4728
```

```
Query: 2314 ctccctgctccctccctccaccggggaagggcatgggctagaagaggaga 2363
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4727 ctccctgctccctccctccaccggggaagggcatgggctagaagaggaga 4678
```

```
 Score =  133 bits (67), Expect = 9e-34
 Identities = 74/75 (98%), Gaps = 1/75 (1%)
 Strand = Plus / Minus
```

```
Query: 2384 aatgttttggctg-cggggtcccccctccattccctggagtttgggggaaggggaatcat
2442
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4657 aatgttttggctggcggggtcccccctccattccctggagtttgggggaaggggaatcat
4598
```

```
Query: 2443 taaagtgctttcaga 2457
            |||||||||||||||
Sbjct: 4597 taaagtgctttcaga 4583
```

```
 Score =  103 bits (52), Expect = 8e-25
 Identities = 52/52 (100%)
 Strand = Plus / Minus
```

```
Query: 2175 agctggttgaattgtctttattaacaaacaggatatccaaggccactacatt 2226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4863 agctggttgaattgtctttattaacaaacaggatatccaaggccactacatt 4812
```

Fig. 5e

Mus_AChR   Mus_AChR   1590   Anti_Mus_AChR   2227   OL: 672   OF1:
934   OF2: 506   .

 Score = 1221 (337.4 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 245/246 (99%), Positives = 245/246 (99%), Strands Minus /
 Plus

Query: 1587 AGTTCACAAACCAGATTTATTGTCAGCGGCCTGTTTTCAAAATCTCTTTCTTGGGGGGTG
1528

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 506 AGTTCACAAACCAGATTTATTGTCAGCGGCCTGTTTTCAAAATCTCTTTCTTGGGGGGTG
565

Query: 1527 GGGGAGAGGTGGGTGCCAGTGCAGGCTCATGGTTGGATGCACGGTGGGTAAGGGAGATCA
1468

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 566 GGGGAGAGGTGGGTGCCAGTGCAGGCTCATGGTTGGATGCACGGTGGGTAAGGGAGATCA
625

Query: 1467 GGAACTTGGTTGAAGTAACCCCCAAGGAAGATGAGAGTAGAACCAACGCTGAAGAGCACC
1408

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 626 GGAACTTGGTTGAAGTAACCCCCAAGGAAGATGAGAGTAGAACCAACGCTGAAGAGCACC
685

Query: 1407 AAAGCTGCCCAAAAACAGACATTGTCCAGGGCCTTCCCCATACGCACCCAGTCGGACAGT
1348

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 686 AAAGCTGCCCAAAAACAGACATTGTCCAGGGCCTTCCCCATACGCACCCAGTCGGACAGT
745

Query: 1347 TCCTCT 1342
             |||| |
Sbjct: 746 TCCTGT 751

 Score = 954 (263.6 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 198/207 (95%), Positives = 198/207 (95%), Strands Minus /
 Plus

Query: 1249 GCAGAGGGCTGCGGTCCAAGTTCCGTGCCGATGCCTCTGACCCTCAAACACGAGTTCGCT
1190

             |||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1007 GCAGTGCTTACCGGTCCAAGTTCCGTGCCGATGCCTCTGACCCTCAAACACGAGTTCGCT
1066

Query: 1189 CCGCGGCTTTTTCAAGATGAGCTCCTCCGCTCTGAGCAGAATGCCCACAGACGAGGCACG
1130

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1067 CCGCGGCTTTTTCAAGATGAGCTCCTCCGCTCTGAGCAGAATGCCCACAGACGAGGCACG
1126

Query: 1129 CCTCGCTGGTGAGGCAGTTCGGGGATCCTCTGGGGGTGGGCTCGAGCCCAGGAGACGCGG
1070

             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1127 CCTCGCTGGTGAGGCAGTTCGGGGATCCTCTGGGGGTGGGCTCGAGCCCAGGAGACGCGG
1186

Query: 1069 CAGCAGCTCTAATAAAATCTGGCGCAG 1043   **Fig. 5f**
             ||||||||||||||||||||| || |

Sbjct: 1187 CAGCAGCTCTAATAAAATCTGCAGCCG 1213

 Score = 591 (163.3 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 119/120 (99%), Positives = 119/120 (99%),   Strands Minus /
Plus

Query: 1053 ATCTGGCGCAGCCGAGGGGATGTAGCATGAGTCGTTGGCGTCCTCAAAGATACGTTGAGC
994
             | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1277 ACCTGGCGCAGCCGAGGGGATGTAGCATGAGTCGTTGGCGTCCTCAAAGATACGTTGAGC
1336

Query: 993 ACGATGACGCAATTCATGACAATGAGCGTGGCAACCACCATGACGAATATAAGATACCTG
934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1337 ACGATGACGCAATTCATGACAATGAGCGTGGCAACCACCATGACGAATATAAGATACCTG
1396

 Score = 546 (150.9 bits), Expect = 3.7e-254, Sum P(4) = 3.7e-254
 Identities = 110/111 (99%), Positives = 110/111 (99%),   Strands Minus /
Plus

Query: 1346 CCTCTCCAGTGGCTTCCTGGTCTCTTGTGCTCTCAGCCACAAAGTTCACAGCATCCACAC
1287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 826 CCTCTCCAGTGGCTTCCTGGTCTCTTGTGCTCTCAGCCACAAAGTTCACAGCATCCACAC
885

Query: 1286 AGCAGCGGATTTCTGGGGCTGCAGCACCCAGGTTCTGGCAGAGGGCTGCGG 1236
             |||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct: 886 AGCAGCGGATTTCTGGGGCTGCAGCACCCAGGTTCTGGCAGAGGGCTGCTG 936

Fig. 5f continued

```
CyclinE2    CyclinE2      2714      Anti_CyclinE2            6773      OL: 1855
OF1: 865          OF2: 2006
```

```
 Score = 7885 (2178.8 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 1577/1577 (100%), Positives = 1577/1577 (100%),   Strands Minus
/ Plus
```

```
Query:  2714 TACAGCTGGCAGCGCAGAGAAGGAAAAAAAAGTTTCTCCAAGCAATGGCAAAACTTTACT
2655
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2006 TACAGCTGGCAGCGCAGAGAAGGAAAAAAAAGTTTCTCCAAGCAATGGCAAAACTTTACT
2065


Query:  2654 TTTAAGCAGTTAAATTTTTTTAACTTTTATTTTTTAAACAATGGGCTAAAAATAAACAGT
2595
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2066 TTTAAGCAGTTAAATTTTTTTAACTTTTATTTTTTAAACAATGGGCTAAAAATAAACAGT
2125


Query:  2594 ATTAAAAGGTTAAGTTTATATAATACATATGTACACAATTAGTGGTGTTTTCTTTTCAGA
2535
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2126 ATTAAAAGGTTAAGTTTATATAATACATATGTACACAATTAGTGGTGTTTTCTTTTCAGA
2185


Query:  2534 CAAAATACTGAAACAAATATTAGTTTAAAAACAAACTATACAGAAGACTTCATACCGTAA
2475
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2186 CAAAATACTGAAACAAATATTAGTTTAAAAACAAACTATACAGAAGACTTCATACCGTAA
2245


Query:  2474 CAATAAATGTATAGTTTCTTCAAAGGGAGAAGAGATTCACATATCTGATAACAAAATAAA
2415
             . ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2246 CAATAAATGTATAGTTTCTTCAAAGGGAGAAGAGATTCACATATCTGATAACAAAATAAA
2305


Query:  2414 CTAGCAATCTAGTTTTCTAATCTACTTTATGAGGCTGGATTTTTTTTTTAGAAAAGCTAA
2355
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2306 CTAGCAATCTAGTTTTCTAATCTACTTTATGAGGCTGGATTTTTTTTTTAGAAAAGCTAA
2365


Query:  2354 TTTAAAATATTTAGAAATAGCTAGCCTATGTACAGCAAGTTTTCATGTCTTTTTTTAATA
2295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2366 TTTAAAATATTTAGAAATAGCTAGCCTATGTACAGCAAGTTTTCATGTCTTTTTTTAATA
2425


Query:  2294 AATAGATTTCTAGGAGTCAGTATATATTTAATACTCTTCTTCCTTAAGAAAATAGAAGTT
2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2426 AATAGATTTCTAGGAGTCAGTATATATTTAATACTCTTCTTCCTTAAGAAAATAGAAGTT
2485


Query:  2234 TAGGTCAAGTGTTAAGCTTTATCACTTTGACACTGTCCTTATCTCACAATGGAGGAATTT
2175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Fig. 5g

```
Sbjct:    2486 TAGGTCAAGTGTTAAGCTTTATCACTTTGACACTGTCCTTATCTCACAATGGAGGAATTT
2545

Query:    2174 AGAAAGGACCTTAACAGTTTCACAAACATAAATAAAGCCTTAGTCACACTAAATTAAAAA
2115
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2546 AGAAAGGACCTTAACAGTTTCACAAACATAAATAAAGCCTTAGTCACACTAAATTAAAAA
2605

Query:    2114 AAAAAATTCCTTAGGGATATCTTAGAGTAGTAAAGTGACTTCCTCATATAAATAGTTTGA
2055
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2606 AAAAAATTCCTTAGGGATATCTTAGAGTAGTAAAGTGACTTCCTCATATAAATAGTTTGA
2665

Query:    2054 AAGGGTACTTAAGTTTTTCACCCAAATTGTGATATACAAAAAGGTTATTACCAAGCAACC
1995
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2666 AAGGGTACTTAAGTTTTTCACCCAAATTGTGATATACAAAAAGGTTATTACCAAGCAACC
2725

Query:    1994 TACATGTCAAGAAAGCCCCAGTTAGGAAGGAGCCACAGCATTTATCTTGTTTATAATTTC
1935
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2726 TACATGTCAAGAAAGCCCCAGTTAGGAAGGAGCCACAGCATTTATCTTGTTTATAATTTC
2785

Query:    1934 TTTGGTACTCCCACTGTTTAGAGCACAGGTTGAACACCATGTTCATCTAAGCCTTATTAG
1875
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2786 TTTGGTACTCCCACTGTTTAGAGCACAGGTTGAACACCATGTTCATCTAAGCCTTATTAG
2845

Query:    1874 TTAAAAAATGTGTTATGGCAAGGCAAATAAACTAGTTTAAAAAACATTAAATTTCACCAT
1815
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2846 TTAAAAAATGTGTTATGGCAAGGCAAATAAACTAGTTTAAAAAACATTAAATTTCACCAT
2905

Query:    1814 TTGTAGAAATTCAAGTTTTATAATAGCTTGCTATAGCAGCTATAGATAAATTAGTCACCT
1755
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2906 TTGTAGAAATTCAAGTTTTATAATAGCTTGCTATAGCAGCTATAGATAAATTAGTCACCT
2965

Query:    1754 TATTACAAAACTAAACCTTTGTAAACAAGTTTAAATTTAATTTTCAAGAACCAAATTGCA
1695
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    2966 TATTACAAAACTAAACCTTTGTAAACAAGTTTAAATTTAATTTTCAAGAACCAAATTGCA
3025

Query:    1694 CTAGTCAAGAGTGTAGGAATTTTGAGAATCTAACAACTAGATTCAAAGTACTGTATCACT
1635
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3026 CTAGTCAAGAGTGTAGGAATTTTGAGAATCTAACAACTAGATTCAAAGTACTGTATCACT
3085

Query:    1634 TAGTATACCCTTTAAGGTAGCACTTATCCAGTCCAAAACTCCAGTGACAAAATTCCTAGT
1575
```

Fig. 5g continued

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3086 TAGTATACCCTTTAAGGTAGCACTTATCCAGTCCAAAACTCCAGTGACAAAATTCCTAGT
3145

Query:    1574 TTATCAAGATAAACACAGTAACACTGGATTAAAGGAAAAACATTGCTATGGTATAGACTG
1515
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3146 TTATCAAGATAAACACAGTAACACTGGATTAAAGGAAAAACATTGCTATGGTATAGACTG
3205

Query:    1514 TGGTTGGCTTCTATCCAGTAACCTTGGGAATGAAGACATCTTTGTAAACAAGTCCTGCTG
1455
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3206 TGGTTGGCTTCTATCCAGTAACCTTGGGAATGAAGACATCTTTGTAAACAAGTCCTGCTG
3265

Query:    1454 TTTCTTTAACAGCTAACATAGGAAATAATTAAATGTATTCTTTAGTGCCAATTGTAAGTT
1395
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3266 TTTCTTTAACAGCTAACATAGGAAATAATTAAATGTATTCTTTAGTGCCAATTGTAAGTT
3325

Query:    1394 TTAAAATCAGAATGGCAGTGTAACTTGTGAATTGGCTAGGGCAATCAATCACAGCACTAC
1335
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3326 TTAAAATCAGAATGGCAGTGTAACTTGTGAATTGGCTAGGGCAATCAATCACAGCACTAC
3385

Query:    1334 TTTCTGTAAAACTTTAGTAGTTCAGTGATACCAGTTCTACCCAATCTTGGTGAATTCCAA
1275
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3386 TTTCTGTAAAACTTTAGTAGTTCAGTGATACCAGTTCTACCCAATCTTGGTGAATTCCAA
3445

Query:    1274 CTTGTTTGCTTAGTTATCTTCTTTAGTGTTTTCCTGGTGGTTTTTCAGTGCTCTTCGGTG
1215
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3446 CTTGTTTGCTTAGTTATCTTCTTTAGTGTTTTCCTGGTGGTTTTTCAGTGCTCTTCGGTG
3505

Query:    1214 GTGTCATAATGCCTCCATTGCACACTGGTGACAACTGTCCCCCTTTTCTGAAGGTGTTTA
1155
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    3506 GTGTCATAATGCCTCCATTGCACACTGGTGACAACTGTCCCCCTTTTCTGAAGGTGTTTA
3565

Query:    1154 TGTAATTTACTTCCTCC 1138
                 |||||||||||||||||
Sbjct:    3566 TGTAATTTACTTCCTCC 3582
```

Score = 805 (222.4 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 161/161 (100%), Positives = 161/161 (100%),   Strands Minus /
Plus

```
Query:    1139 CCAGCATAGCCAAATAGTTTGTATGTGTCTGGATATTATGTCTGTCTTCCATAGGAATCT
1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Fig. 5g continued

Sbjct: 3967 CCAGCATAGCCAAATAGTTTGTATGTGTCTGGATATTATGTCTGTCTTCCATAGGAATCT
4026

Query: 1079 TCTTAAAAGTCTTCAGCTTCACTGGACTAGTACTTTTTACTACATTGACAAAAGGTACCA
1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4027 TCTTAAAAGTCTTCAGCTTCACTGGACTAGTACTTTTTACTACATTGACAAAAGGTACCA
4086

Query: 1019 TCCAATCTACACATTCTGAAATACTGTCCCACTCCAAACCT 979
             |||||||||||||||||||||||||||||||||||||||||
Sbjct: 4087 TCCAATCTACACATTCTGAAATACTGTCCCACTCCAAACCT 4127

 Score = 581 (160.5 bits), Expect = 0.0, Sum P(4) = 0.0
 Identities = 117/118 (99%), Positives = 117/118 (99%), Strands Minus /
Plus

Query: 982 ACCTGAGGCTTTCTTAACCACTTCAATGGAGGTAAAATGGCACAAGGCAGCAGCAGTCAG
923
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 4615 ACCTGAGGCTTTCTTAACCACTTCAATGGAGGTAAAATGGCACAAGGCAGCAGCAGTCAG
4674

Query: 922 TATTCTGTACTGGAACTCTAATGAATCAATGGCTAGAATACACAGATCTAAAAGCTGA 865
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct: 4675 TATTCTGTACTGGAACTCTAATGAATCAATGGCTAGAATACACAGATCTAAAAGCTAA 4732

Fig. 5g continued
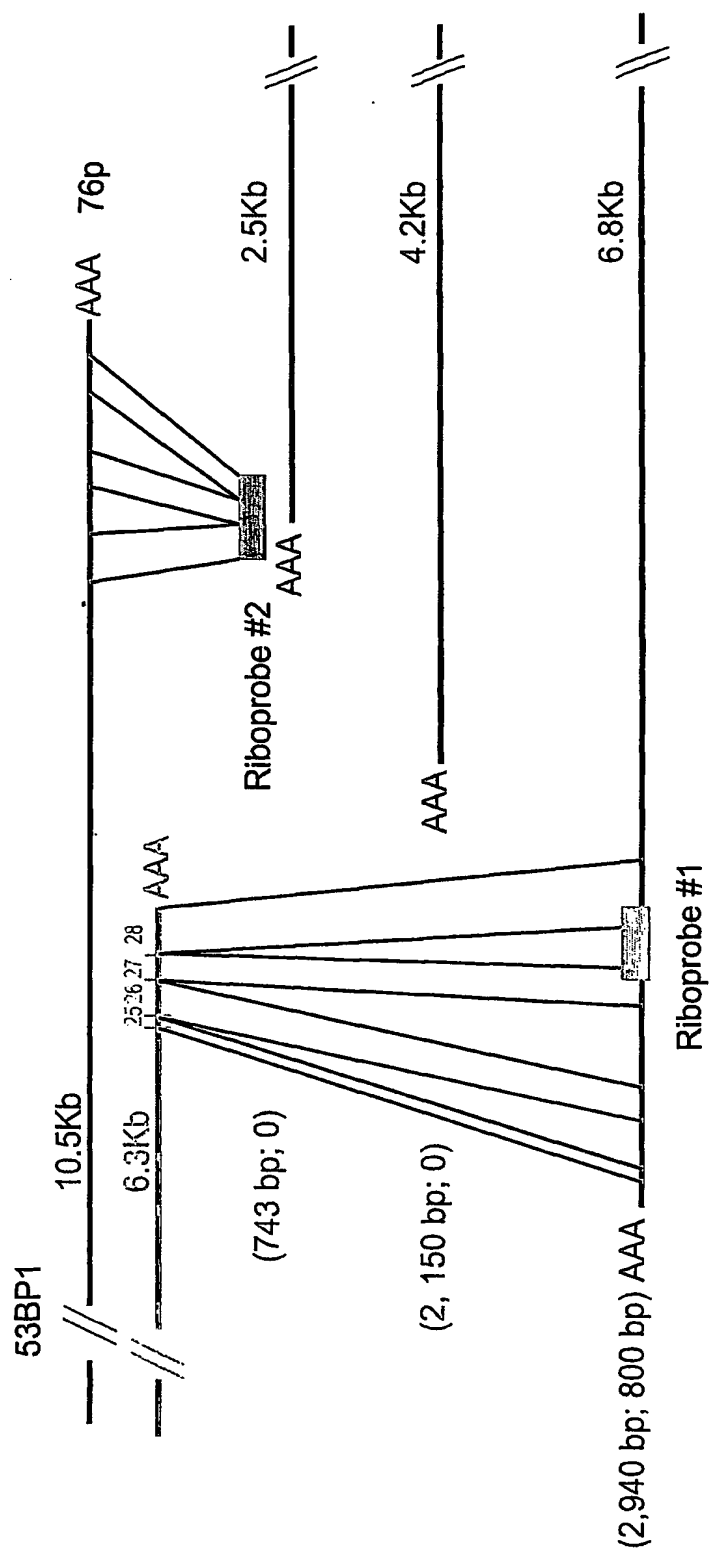
Fig. 6
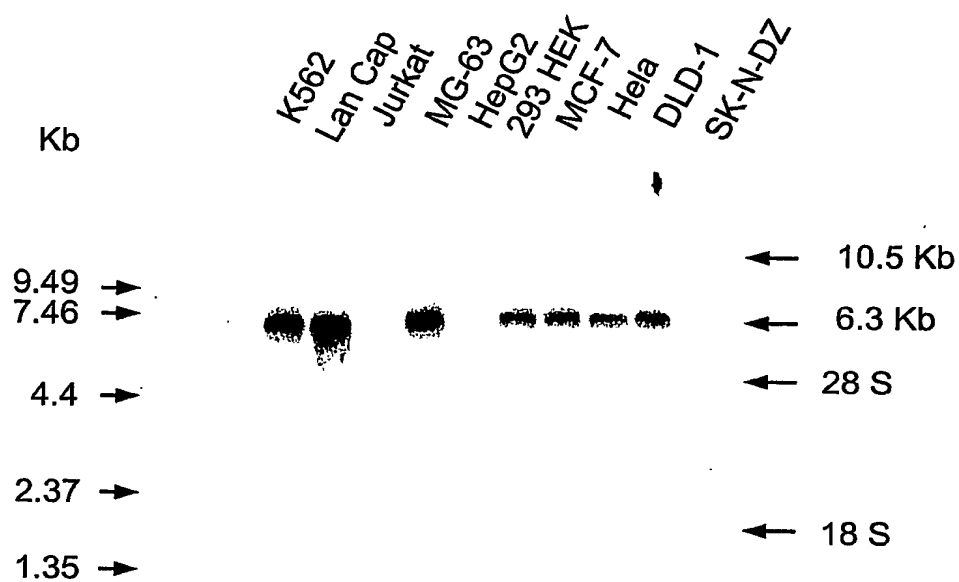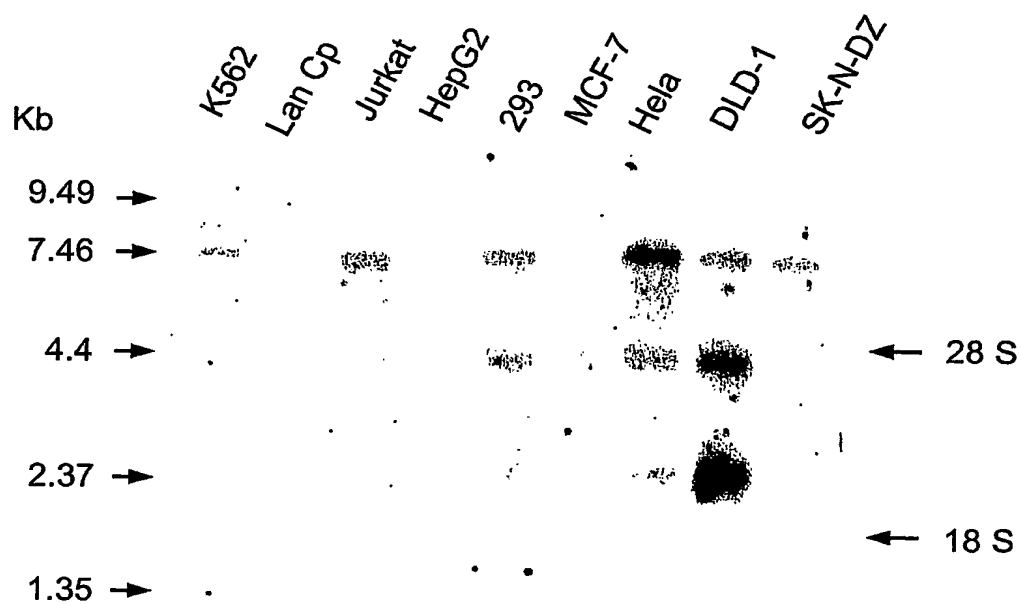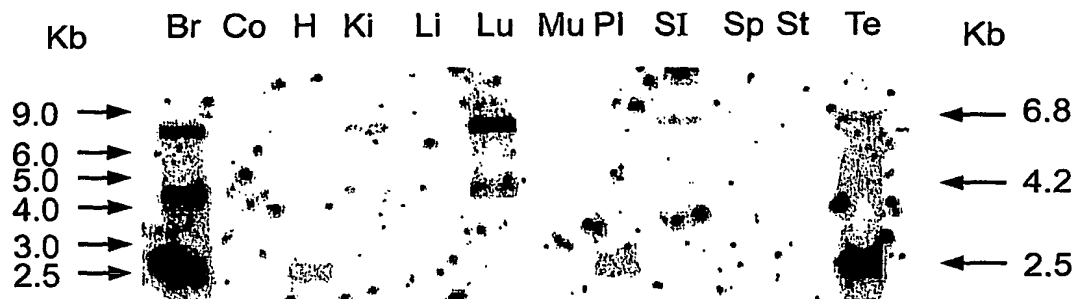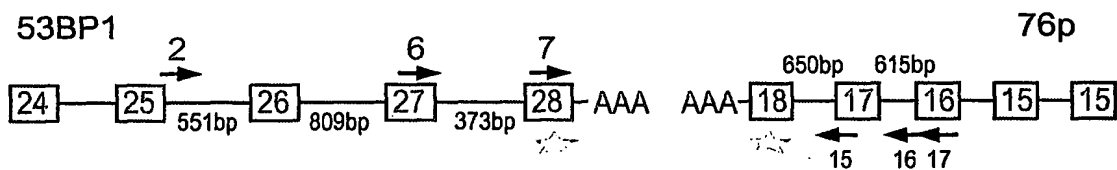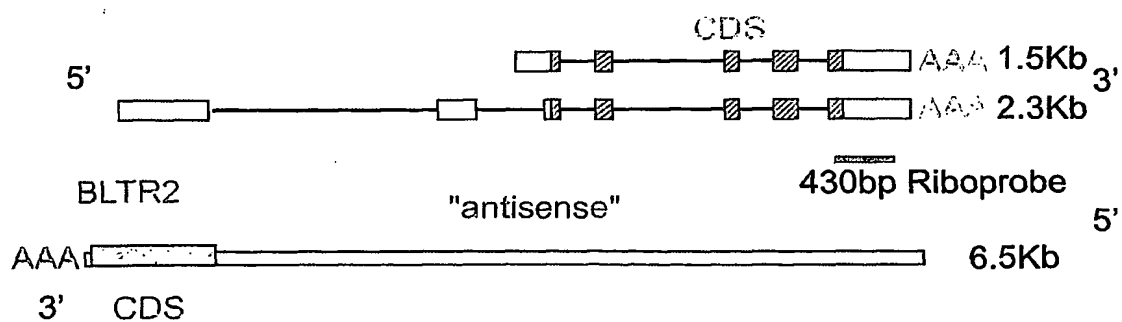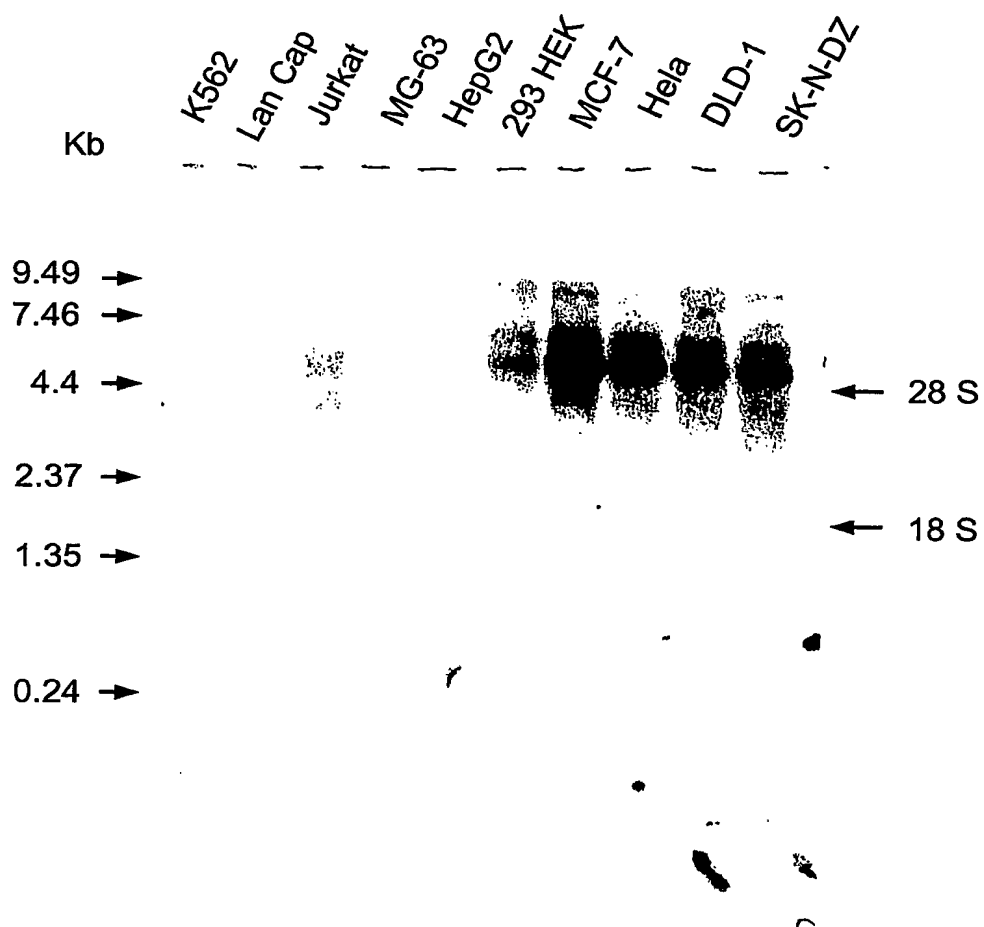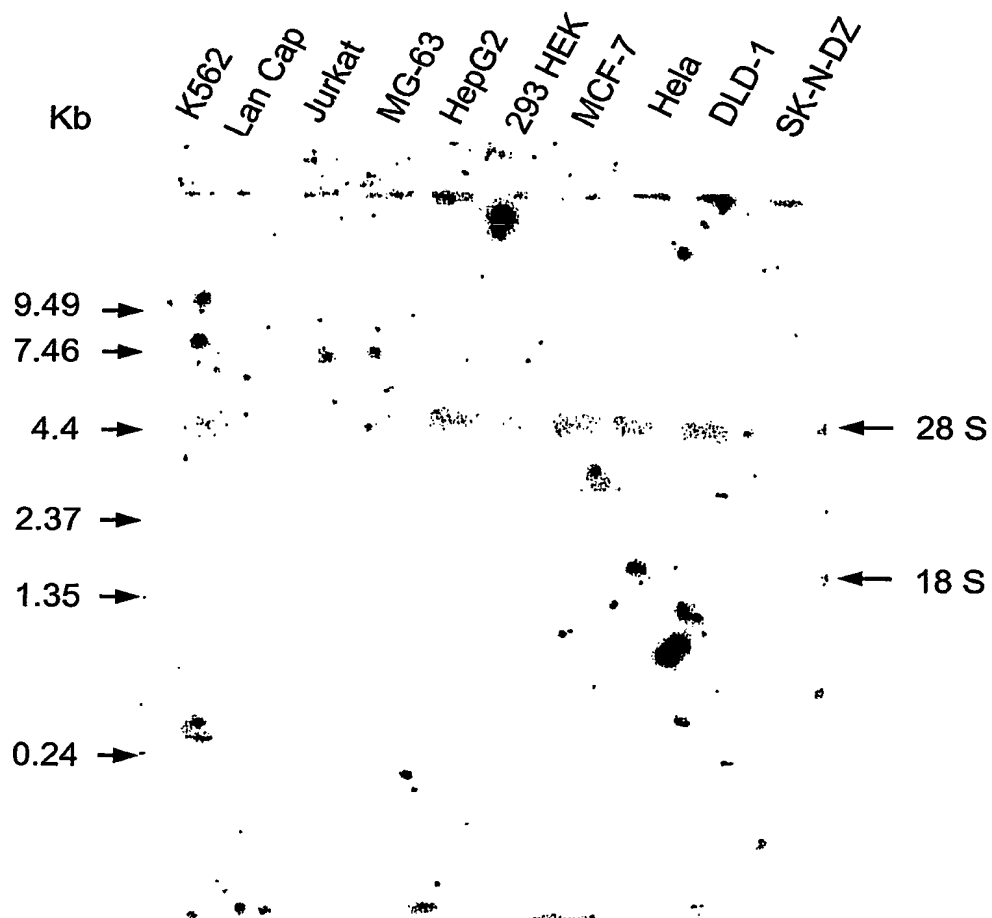
Fig. 7

Fig. 8

Fig. 9

Fig. 10

Fig. 11

Fig. 12

Fig. 13

Fig. 14

Fig. 15b

With Apaf-1 specific riboprobe

Fig. 15a

With EB-1 specific riboprobe
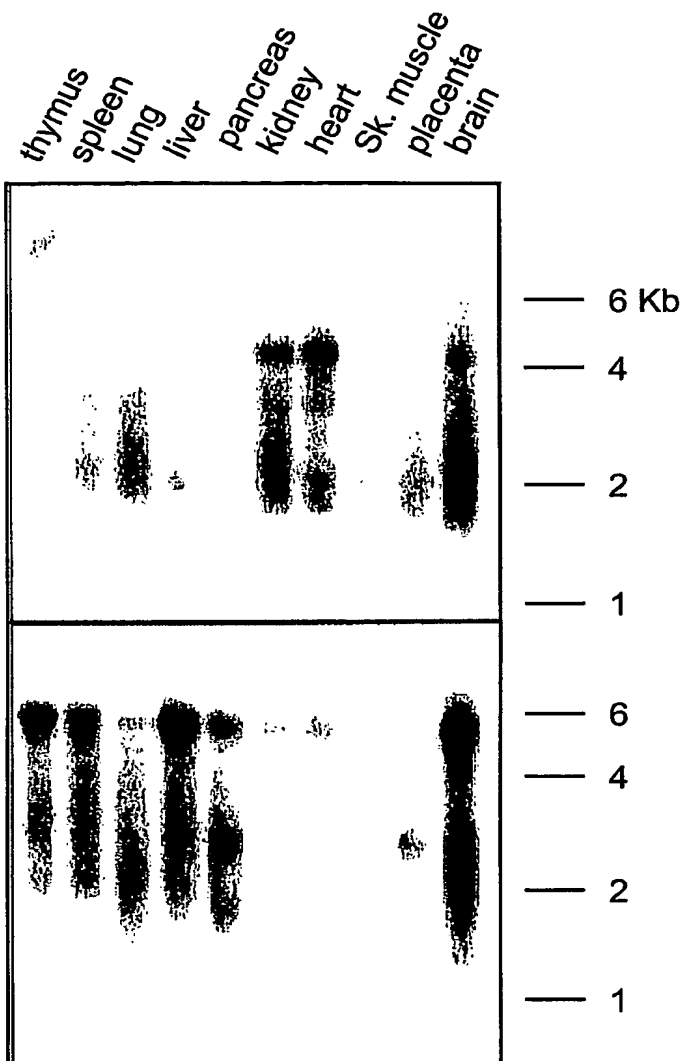
Fig. 16

Fig. 17a

AChR-ε
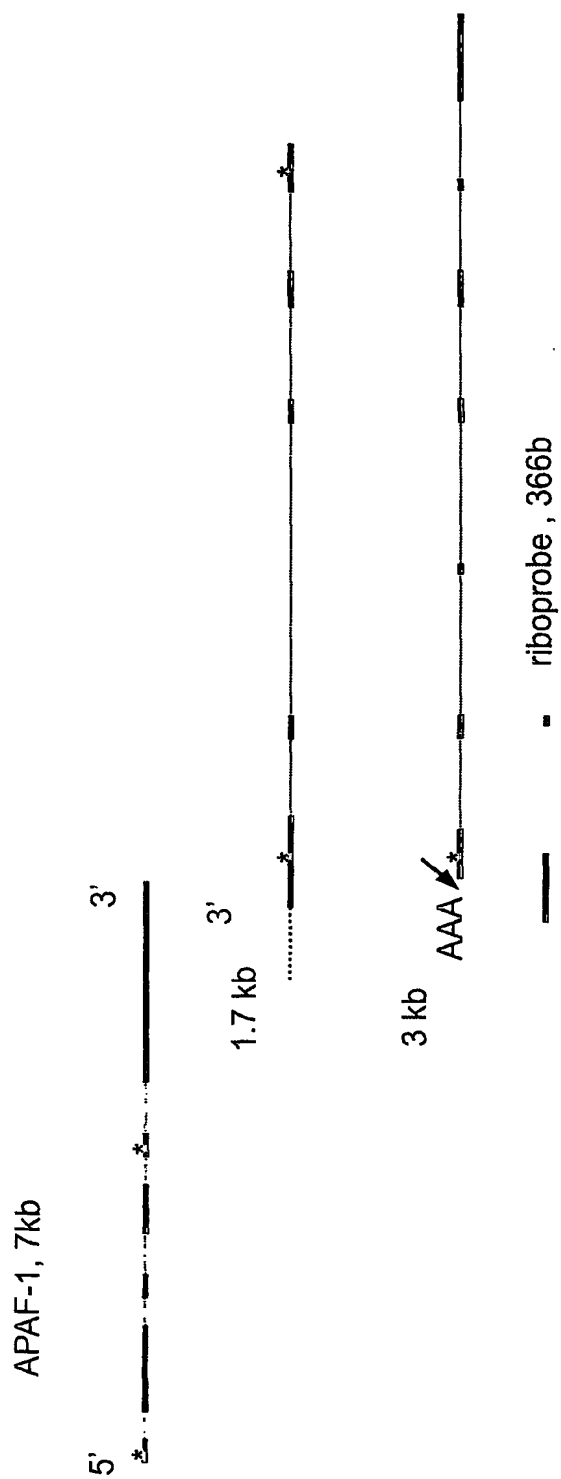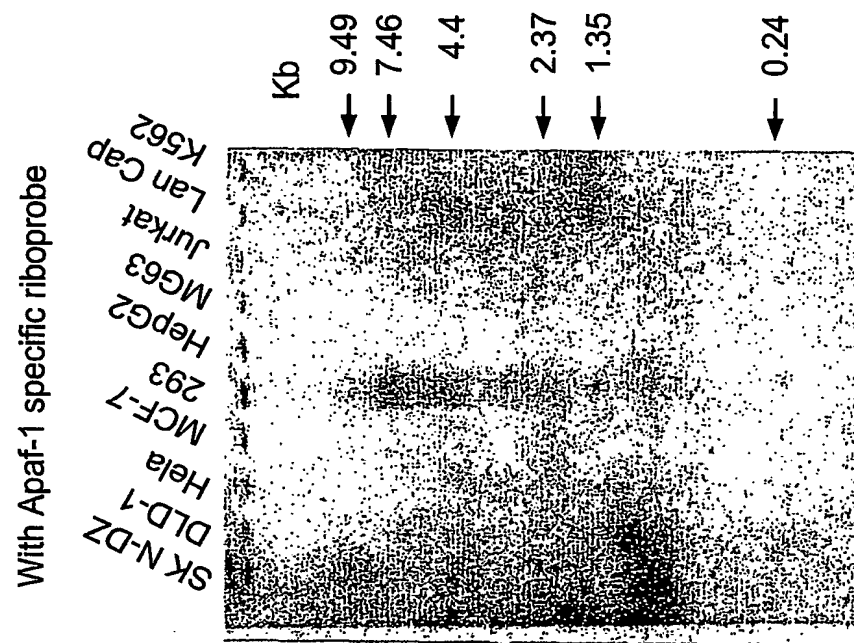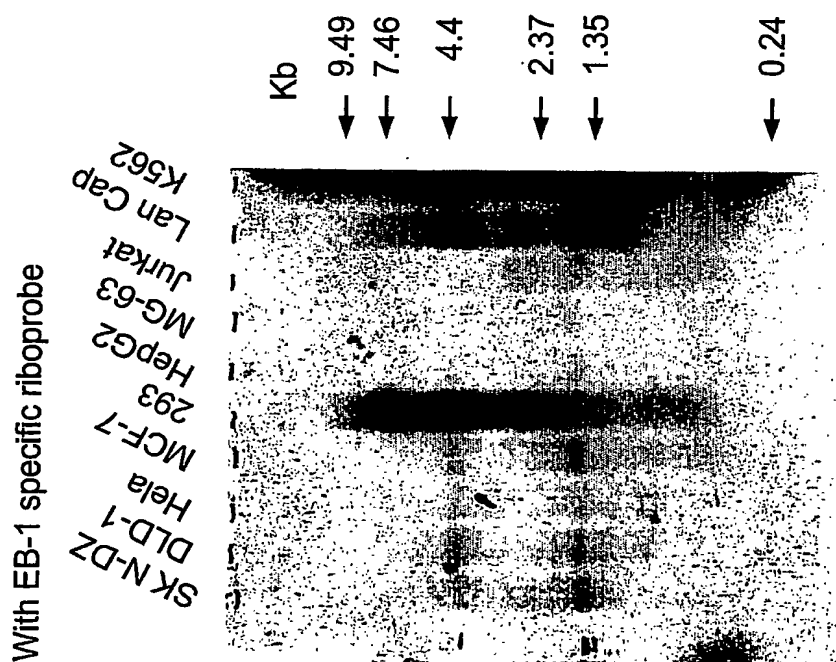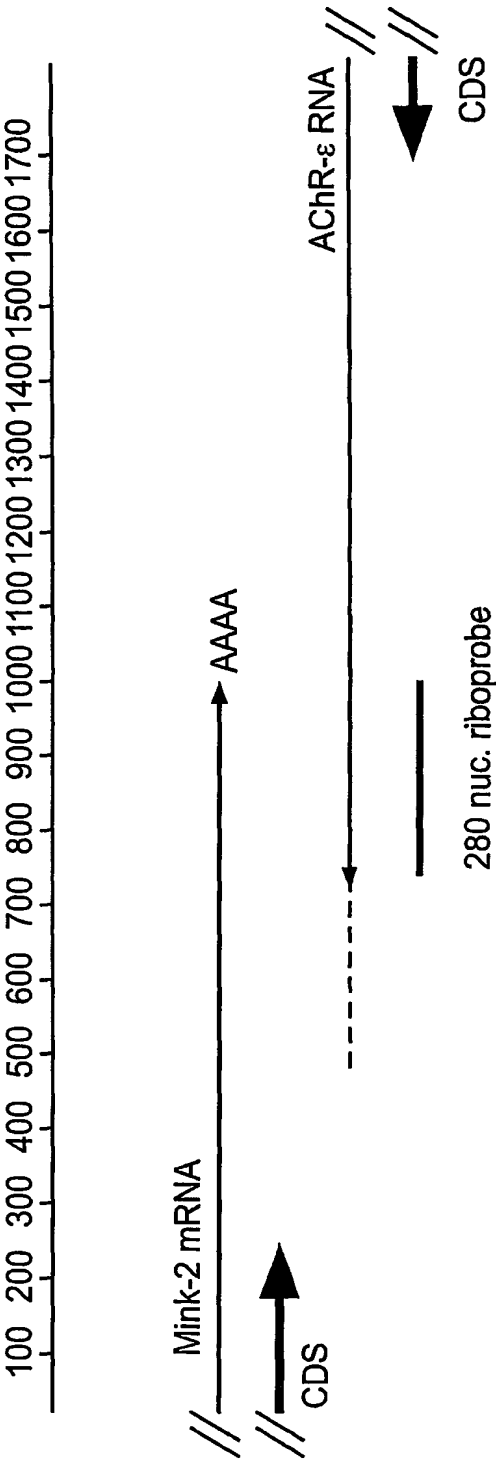
Fig. 17b

Mink
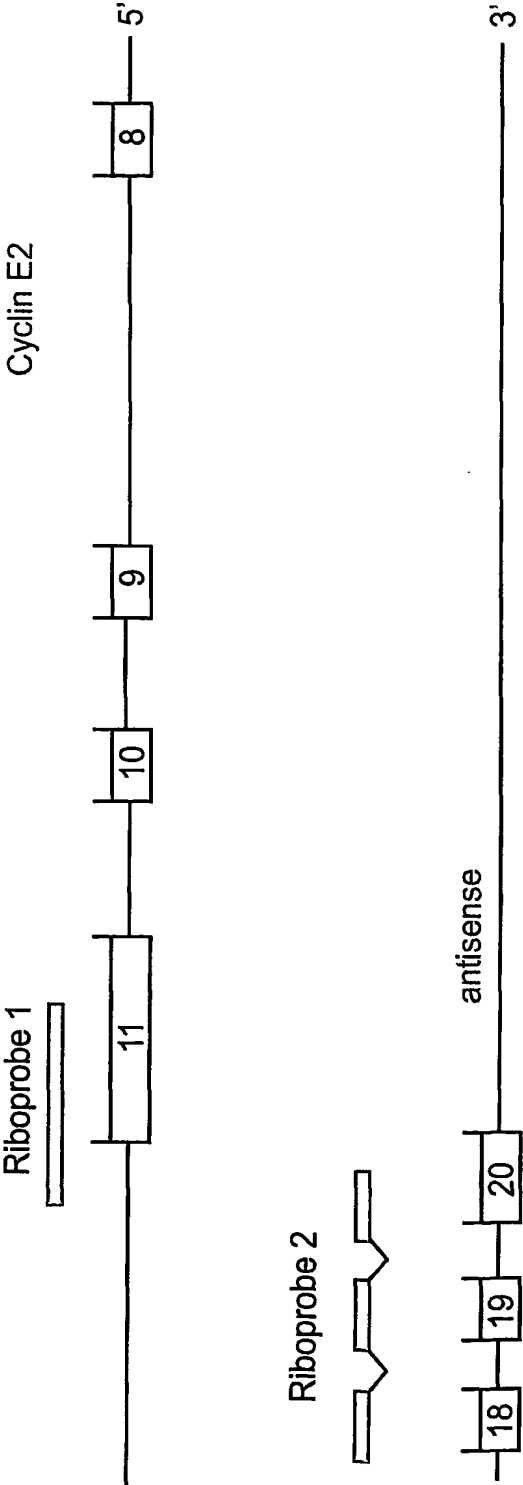
Fig. 18
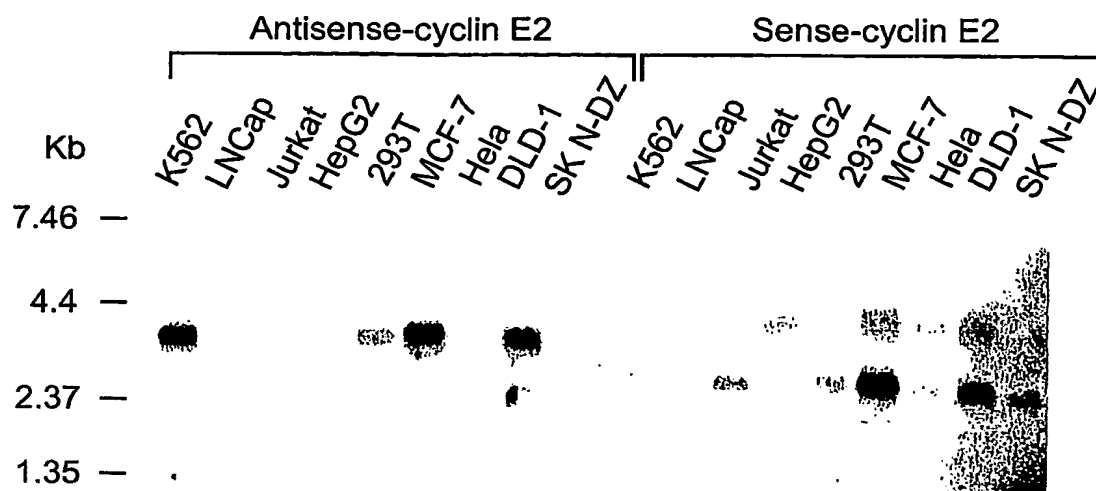
Fig. 19a   Fig. 19b

Fig. 20
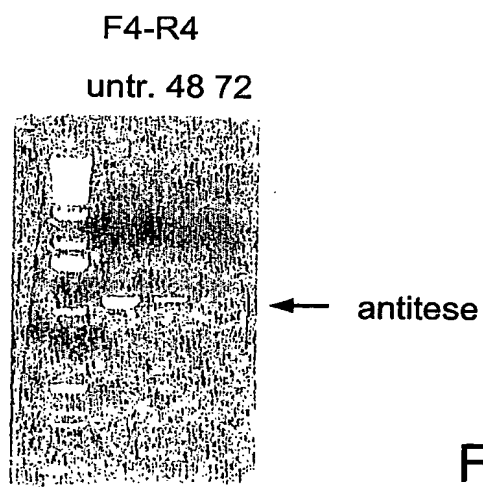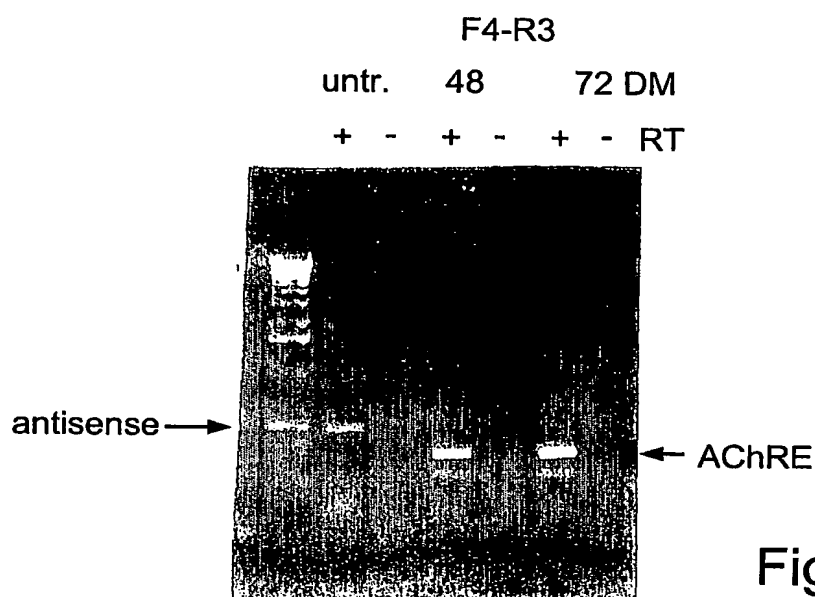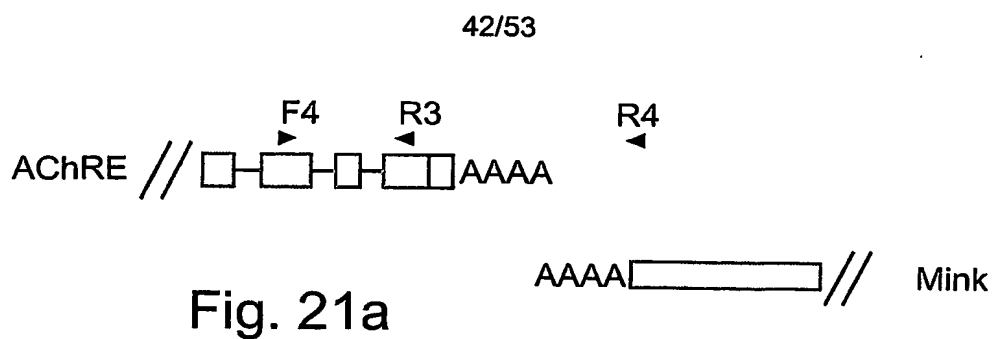
Fig. 21a

Fig. 21b

Fig. 21c
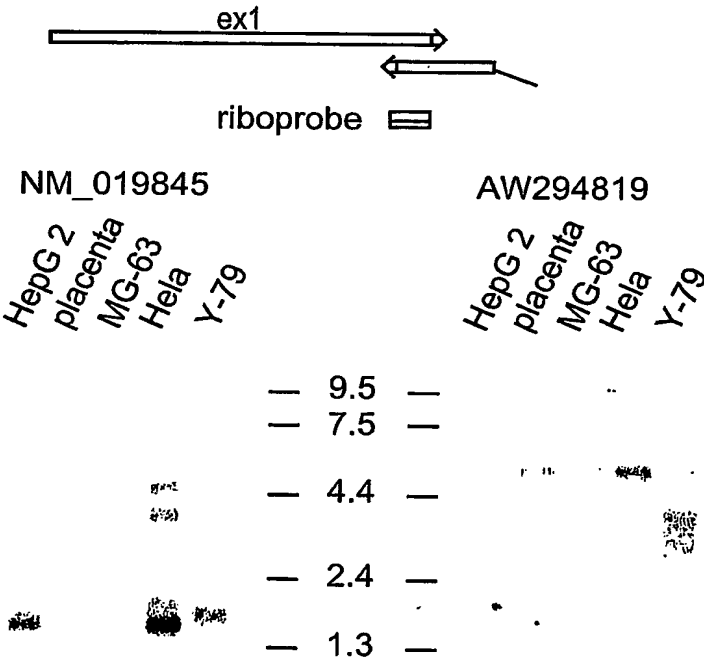
Fig. 22a

Fig. 22b
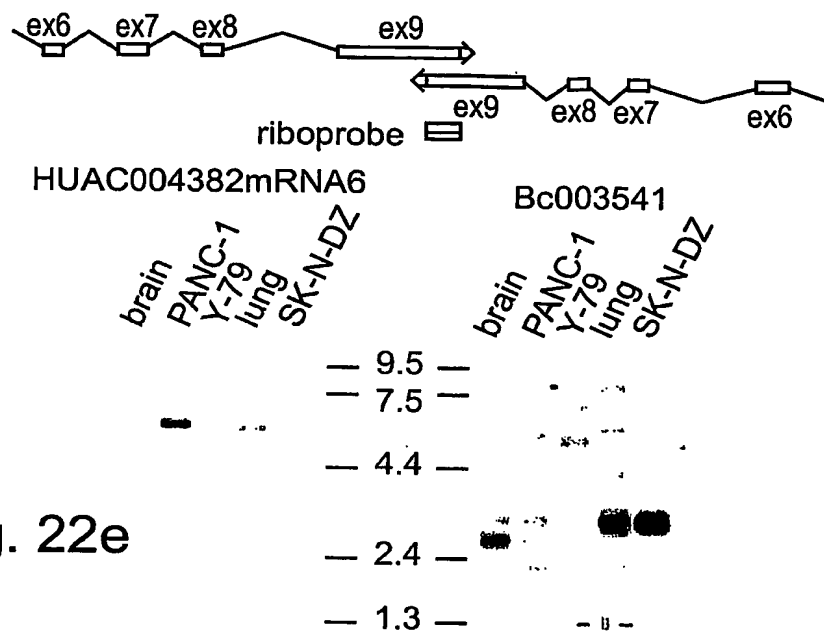
Fig. 22c
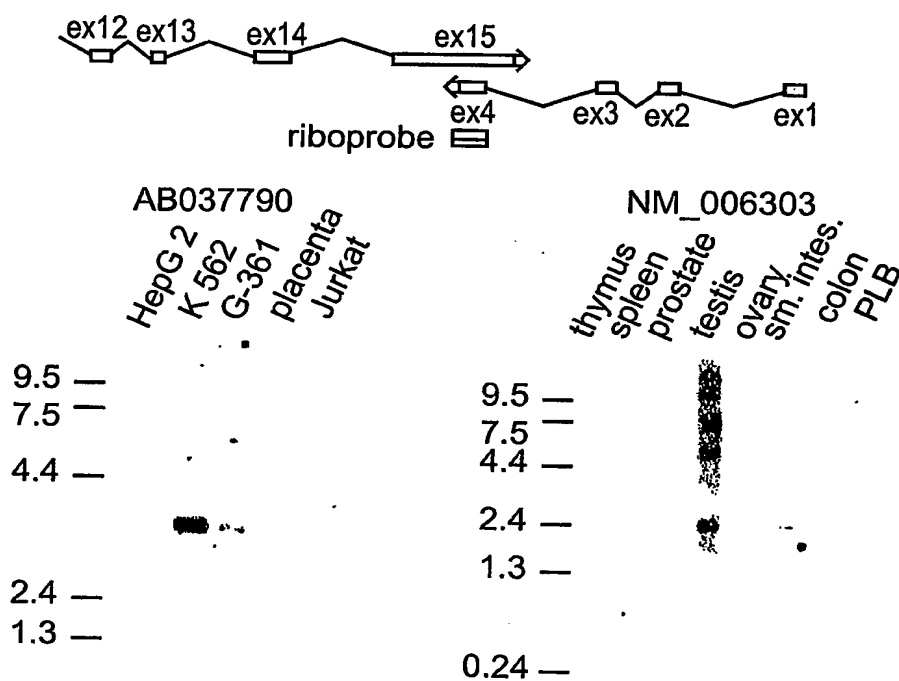
Fig. 22d
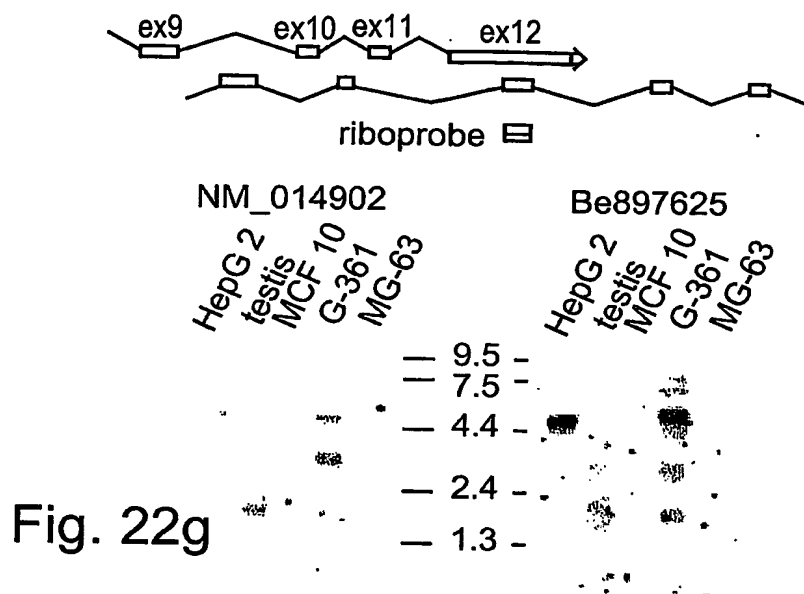
Fig. 22e
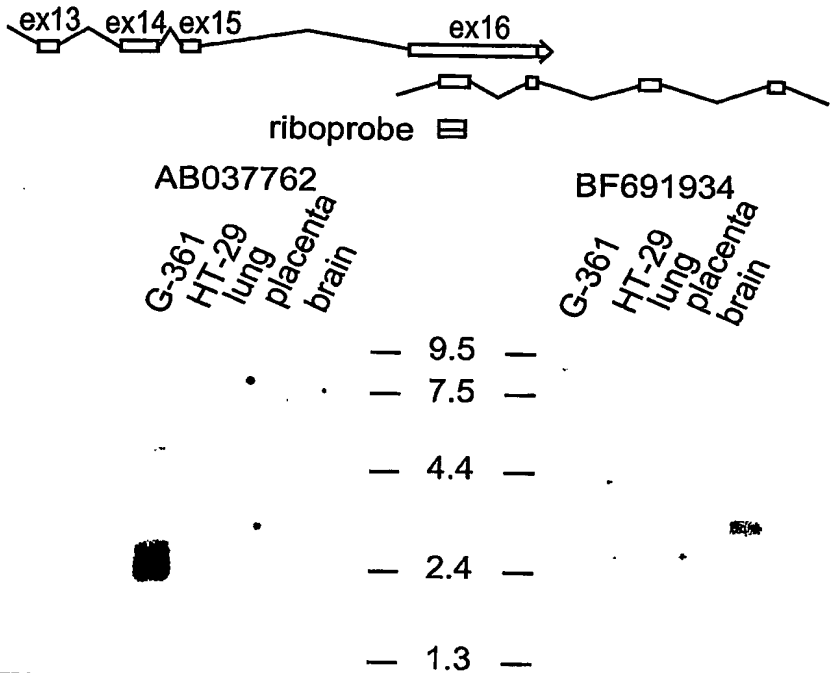
Fig. 22f
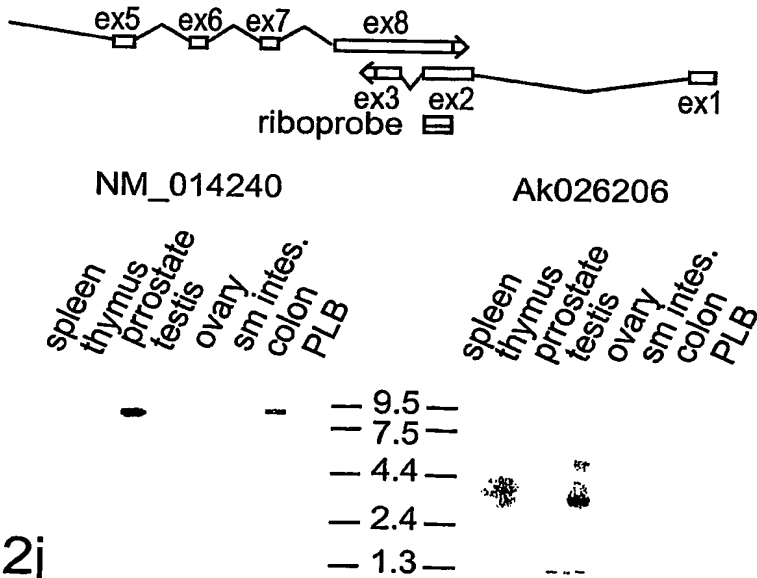
Fig. 22g

Fig. 22h

Fig. 22i

Fig. 22j

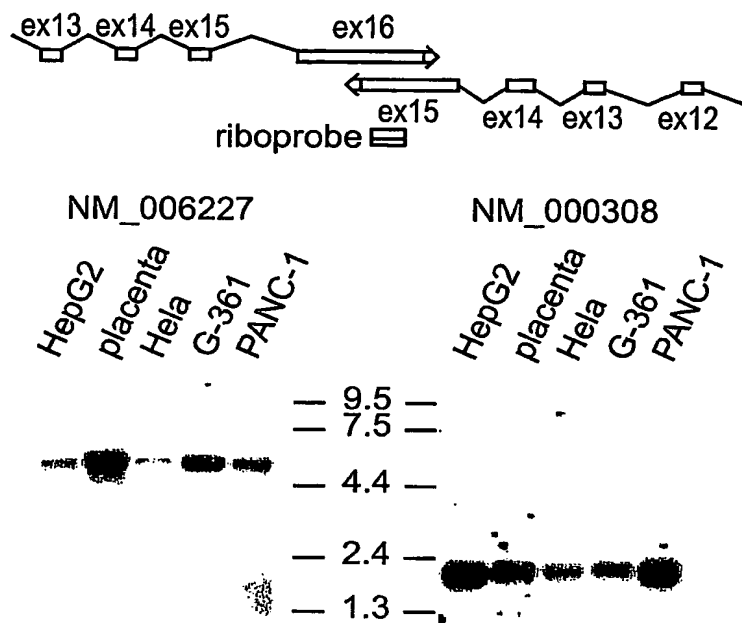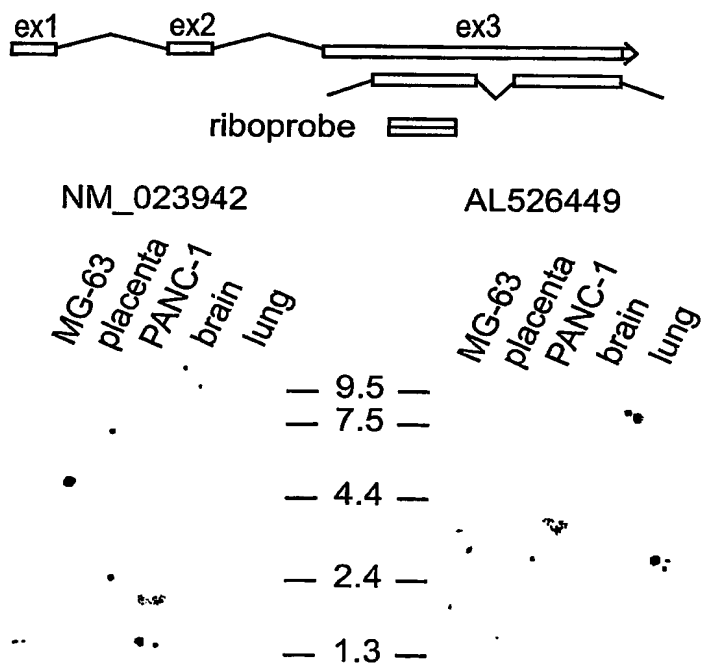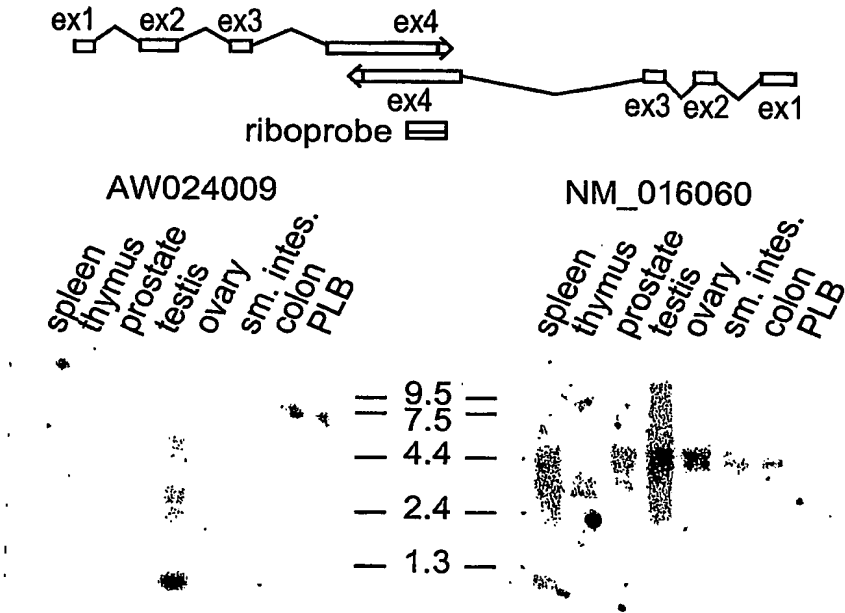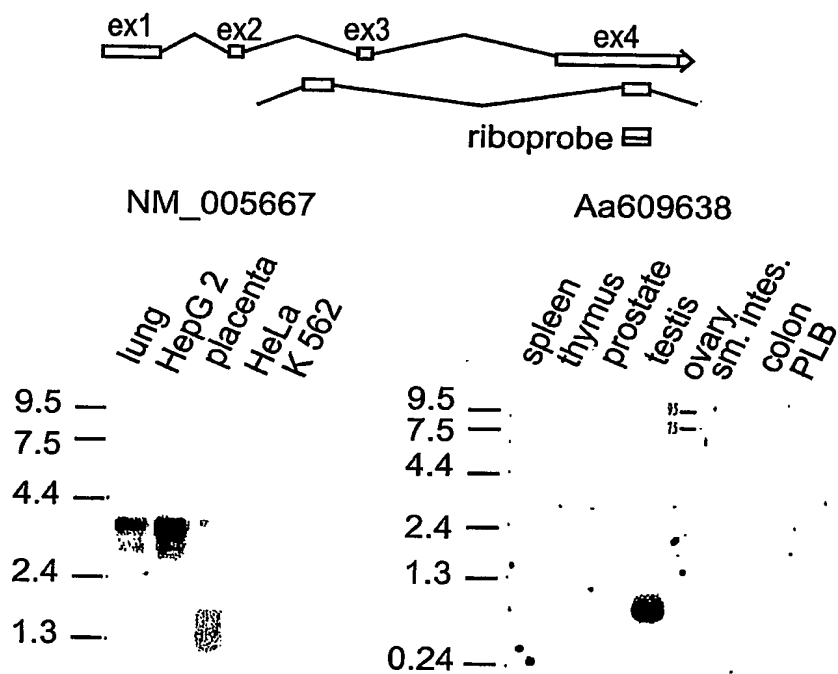| Probe	Probe pass thresh hold	Anti sense pair ver.	Actu al pool	Virtu al pool	Y79 28	Y79 52b	ES-2 29	ES-2 53	HEK-293 30b	HEK-293 78	K-562 31b	K-562 55b	HT-29 33b	HT-29 56	SK-N DZ 34b	SK-N DZ 89	MG-63 35	MG-63 58	SNUI 36b	SNUI 79	NI564 45	NI564 63	HeLa 46b	HeLa 64	MCF10 50b
Actin 1uM	+	+ CTRL	472	442	0.56	0.71	0.89	0.89	0.62	0.95	0.71	0.73	0.41	0.47	0.66	0.62	0.78	0.8	1.01	1.12	2.24	2.6	0.91	1.09	0.92
GAPDH 30um	+	+ CTRL	1847	1710	0.97	0.96	0.82	0.78	0.53	0.51	0.97	0.86	0.47	0.52	0.87	0.73	1.03	0.8	1.88	1.65	1.28	0.87	1.38	1.33	0.84
HUMM HBA123	+	+ CTRL	2245	2282	1.43	1.42	0.85	0.78	1.16	1.06	1.26	1.02	0.82	0.88	1.1	0.95	1.95	1.28	0.88	0.75	0.91	0.54	1.37	1.25	1.14
Known 10AC	+	VER.	378	426	1.82	1.49	0.65	0.54	0.94	0.91	1.32	0.82	1.03	0.66	1.61	1.57	2.34	2.37	0.8	0.65	1.06	0.55	1.15	0.9	0.79
Known 10C	+	VER.	95	103	1.04	1.27	0.78	0.84	-1	1.51	1	-1	1	1.16	-1	1.78	1	1	0.74	1.01	0.79	1	1.49	1	0.58
Known 9AC	+	VER.	852	959	2	1.84	0.81	0.69	1.07	1.26	1.56	1.14	1.11	0.85	1.55	1.27	1.14	0.83	0.76	0.63	0.89	0.49	1.28	1.25	0.77
260AC	+		154	197	1.34	1.95	0.35	1	0.9	1.79	0.84	1	0.71	1.42	0.87	1.35	0.47	1	0.59	1.3	0.78	1	1.14	1	1.31
260C	-		33	48	-1	1	0.93	1	-1	1.8	-1	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1
261AC	+	VER.	344	357	1.28	1.15	0.46	0.65	1.34	1.22	1.76	1.09	1	0.69	1.65	1	1.04	0.71	0.86	0.64	1.32	0.39	2.09	1.01	0.6
72C	-	IND. VER.	63	69	1	1	0.78	1	1	1	-1	1	1	2.32	1.14	1.78	-1	-1	1	1	1	1	1	1	1

Fig. 23

	Lung		Testis		Placenta		DU 145		T24		G-361		Jurkat		HepG2		MCF7		PANC-1	H1299	Brain
68b	70	74	71	75	72b	76	82	83	69	51	80	60	88	66	86b	67b	85	84	37	43	77
0.98	1.34	1.37	0.45	0.55	1.51	1	1.19	1.1	1.52	0.94	1.41	1.36	1.33	1	0.34	0.55	1.32	1.13	0.86	0.9	1.16
0.82	0.37	0.31	0.36	0.28	0.55	0.44	0.93	0.69	0.96	0.57	1.4	1.19	1.36	0.91	0.85	0.81	0.78	0.5	0.62	2.16	0.86
1.04	1.44	0.9	1.16	0.71	1.19	0.9	1.12	0.82	0.97	0.55	0.95	0.71	1.13	0.62	1.37	1.32	1.31	0.83	1.13	1.65	0.66
0.78	1.74	1.22	1.87	1.34	1.42	1.13	1.17	0.95	0.77	0.35	1.42	1	0.94	0.84	1.12	1.16	0.96	0.59	1.76	1.48	1.43
1.61	1.59	1	0.89	1	1	1	1	1	1	0.57	1	1	1	1	0.85	1.3	1	1	1.65	1.55	1
0.84	2.11	1.27	1.54	1.16	1.04	0.67	1.05	0.9	0.86	0.52	1.38	0.79	1.34	1.1	1.03	1.09	1.48	1.11	1.52	1.29	1.2
3.64	1.01	2.58	1.43	3.45	1.52	3.07	1.02	1	0.56	0.72	0.84	0.5	1	1	0.84	1.55	1.1	1.09	1	1.04	4.89
1	1	1	1	1	1	2.46	1	-1	1	-1	1	1	1	1	1	1.68	1	1	-1	-1	1
1	1	1	2.17	1.31	1.39	0.83	1.36	0.72	0.96	0.52	1	0.58	1.95	1	1.5	0.94	1.86	1	-1	-1	1
1.05	1	1	-1		1	1	1.02	1	1	0.99	1	1	-1	1.83	1	1	1	1	1	1	1

Fig. 23 (Cont.)

50/53

FIG. 24a

2091_0 Z19505_0   5019   M78914_25   801   OL: 187

Query: 222 accccctgccgtagccatcttgcctctctgctgagcggaagcccccgttcggctcctgtc 281
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 160 accccctgccgtagccatcttgcctctctgctgagcggaagcccccgttcggctcctgtc 101

Query: 282 tgttagcggcctctctaggctaccactgacaccgtctctgtggcccggagcctaagagac 341
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 100 tgttagcggcctctctaggctaccactgacaccgtctctgtggcccggagcctaagagac 41

Query: 342 cggaagttcgtgtttccaggcgc 364
           |||||||||||||||||||||||
Sbjct: 40 cggaagttcgtgtttccaggcgc 18

  Score = 87.7 bits (44), Expect = 1e-18
  Identities = 44/44 (100%)
  Strand = Plus / Minus

Query: 183 acctgctagtaggacgcagaaagacagaaggcgaaggagacccc 226
           ||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 229 acctgctagtaggacgcagaaagacagaaggcgaaggagacccc 186
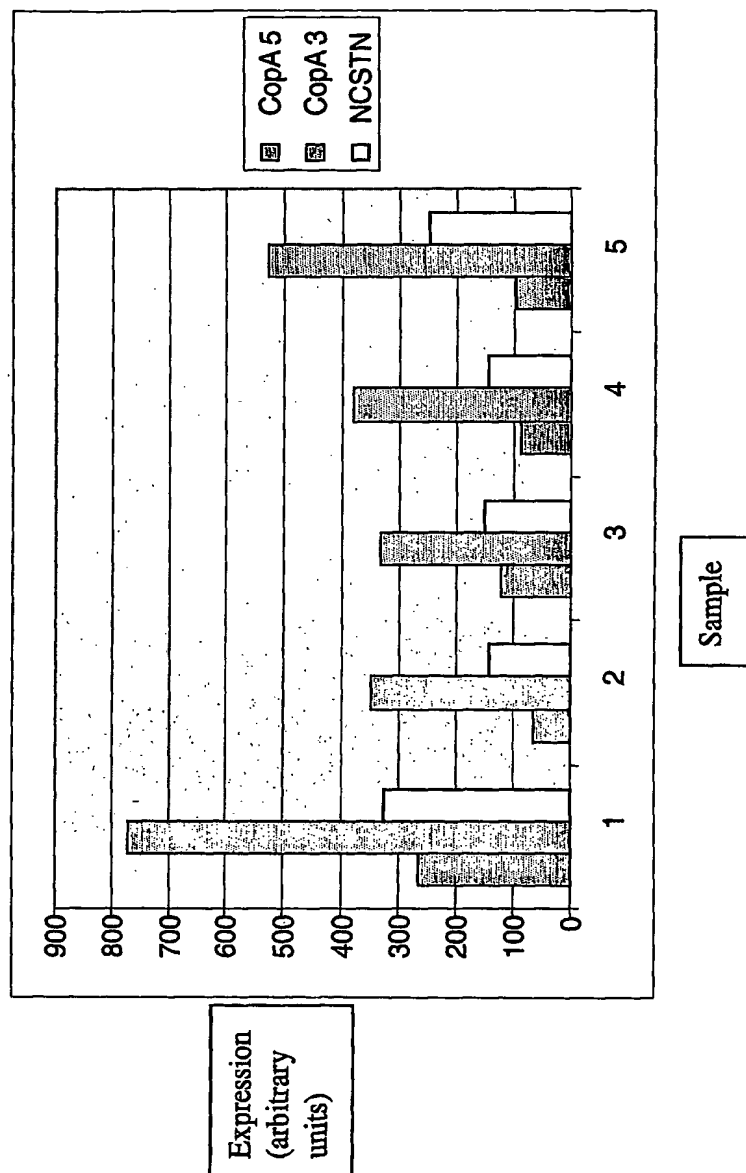
FIG. 24b

FIG. 25a

```
3673_0  T24746_0        3068    N31850_0        1705    OL: 585
Query:  2480  tgagttatgttgtcttttattagcagggaatgtcatcacagattggatagtacatccagg  2539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1705  tgagttatgttgtcttttattagcagggaatgtcatcacagattggatagtacatccagg  1646


Query:  2540  tgcaatgtcaccatcagcaaggtcagcttgacactcaagtggaagattagggaagaatga  2599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1645  tgcaatgtcaccatcagcaaggtcagcttgacactcaagtggaagattagggaagaatga  1586


Query:  2600  ctaggatnnnnnnnnnnnnnggagggcaccaagggaaagggatgatggggtgagctggcgagt  2659
              |||||||            |||||||||||||||||||||||||||||||||||||||||
Sbjct:  1585  ctaggataaaaaaaaaaggagggcaccaagggaaagggatgatggggtgagctggcgagt  1526


Query:  2660  gtgggtgggaaatgaaatgtttattgaggatctgctttgtgctgggcactttaatccaca  2719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1525  gtgggtgggaaatgaaatgtttattgaggatctgctttgtgctgggcactttaatccaca  1466


Query:  2720  ttttatcgtttacttttcaaacagatgcaccttaccccacccccaatgctctgtccctgc  2779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1465  ttttatcgtttacttttcaaacagatgcaccttaccccacccccaatgctctgtccctgc  1406


Query:  2780  agatatcagaagacagtgtgattttcatgctctgaagttcagttttacatccaagcatcc  2839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1405  agatatcagaagacagtgtgattttcatgctctgaagttcagttttacatccaagcatcc  1346


Query:  2840  ctctctgtttttttaacaatccaaagacaggccnnnnnnnngcaccacagtttattaagtac  2899
              |||||||||||||||||||||||||||||||||          ||||||||||||||||||||
Sbjct:  1345  ctctctgtttttttaacaatccaaagacaggccaaaaaaagcaccacagtttattaagtac  1286


Query:  2900  ttactaagcacccatccactgccccacactgtggcaaggattgtgaggggtaaagaagca  2959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1285  ttactaagcacccatccactgccccacactgtggcaaggattgtgaggggtaaagaagca  1226


Query:  2960  tggggcacaatattctgctgccttcatgtaacttacagtctcacaaataaatagaacttc  3019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1225  tggggcacaatattctgctgccttcatgtaacttacagtctcacaaataaatagaacttc  1166


Query:  3020  agttgaaatactgacattaattaaatagagttgtaataaatacaa  3064
              |||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1165  agttgaaatactgacattaattaaatagagttgtaataaatacaa  1121
```

Fig.25b

>45499 T24746_1 (45497 T24746_P2) #GENE_SYMBOL IL24 ;MDA7 ;ST16 #GO_F
#GO_Acc 5125 #GO_Desc cytokine activity #CL 1 #DB sp #EN IL24_HUMAN
#GO_P #GO_Acc 6915 #GO_Desc apoptosis #CL 1 #DB sp #EN IL24_HUMAN #GO_P
#GO_Acc 6955 #GO_Desc immune response #CL 1 #DB sp #EN IL24_HUMAN #GO_C
#GO_Acc 5576 #GO_Desc extracellular #CL 3 #DB PROLOC #EN PROLOC #GO_C
#GO_Acc 5576 #GO_Desc extracellular #CL 1 #DB sp #EN IL24_HUMAN
>45290 W01276_0 (45289 W01276_P1) #GENE_SYMBOL TOSO #TAA blood-tumor
#GO_F #GO_Acc 8189 #GO_Desc apoptosis inhibitor activity #CL 1 #DB sp
#EN O60667 #GO_P #GO_Acc 6916 #GO_Desc anti-apoptosis #CL 1 #DB sp #EN
O60667 #GO_P #GO_Acc 6968 #GO_Desc cellular defense response #CL 1 #DB
sp #EN O60667 #GO_C #GO_Acc 16020 #GO_Desc membrane #CL 3 #DB PROLOC
#EN PROLOC

1
SEQUENCE LISTING

<110> Levanon Erez, et al.

<120> METHODS AND SYSTEMS FOR IDENTIFYING NATURALLY OCCURRING ANTISENSE
TRANSCRIPTS AND METHODS AND KITS UTILIZING SAME

<130> 27987

<150> US 09/718,407
<151> 2000-11-24

<150> US 09/732,938
<151> 2000-12-11

<150> US 09/785,439
<151> 2001-02-20

<150> US 09/907,923
<151> 2001-07-18

<150> US 009/993,398
<151> 2001-11-06

<150> US 10/201,605
<151> 2002-7-24

<160> 44

<170> PatentIn version 3.1

<210> 1
<211> 190
<212> DNA
<213> Homo sapiens

<400> 1
ggacccagga tatgagcgga aaacactttc tctacttaga tacaactttt tcctgtgcgc   60

atgcctgtaa tcccagctac tcaggaggct gaggcaggag aatcccttga acccaggagg   120

2

cagaggttgc ggtgagccaa gatctcacca ttgcactcca gcctgggcaa taagaacaaa   180

actccgtctc   190

<210> 2

<211> 783

<212> DNA

<213> Homo sapiens

<400> 2
gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa   60

gaattcagtc cttgtggatg aactgtaaac agcacccttc ctctaagatg ccgaagatca   120

tagtttgtgg tttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc   180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc   240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa   300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa   360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta   420

atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca   480

gtgaaagatg aacactgggc catttttaaga tggaaacaag gtgggggttg gatagagagt   540

tatatgcagc ctcttttgca cctcgttggt atttgtaaga ccacattttt ttctccctag   600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag   660

tttcagattg cagtacagct atgtaactga gtaagacagg gagaaatatt aatccgtgag   720

tggctcccag taagaccatg gccaaataca tcctgaagta gaatatctgg aaaatttgag   780

att   783

<210> 3

<211> 1649

<212> DNA

<213> Homo sapiens

<400> 3
gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa   60

gaattcagtc cttgtggatg aactgtaaac agcacccttc ctctaagatg ccgaagatca   120

tagtttgtgg tttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc   180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc   240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa   300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa   360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta   420

3

```
atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca    480

gtgaaagatg aacactgggc cattttaaga tggaaacaag gtggggttg  gatagagagt    540

tatatgcagc ctcttttgca cctcgttggt atttgtaaga ccacattttt ttctccctag    600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag    660

tttcagattg cagtacagcc ttcaaatcat ctgggcccaa gttaaaacag aaggaattta    720

aaaaaaaaac acagtcactg tcttagaaga tgactcatat gctaagacag gtctgcctcc    780

ctgactcaga atgctgagtg actcctgaca ttattagttg gaatgggaag tgtaaggtca    840

agttggggtc tttacctgca tgacgaaacc acttcttgta atgacagact tttactgtgt    900

tggttagaat agccagtcct tggggagcct ctagtctgtt gtagctgaat gatttggaag    960

tgttctttca ctttttactt ttgtcctcag cattacctac atgaacttgt tatcaaaacc    1020

cttgtccagc acaacctctt ttatatgctg catcagttcc tgcagtacca cgtcctcagc    1080

gactccaaac ctttggcttg tctgctgtta tccctagaga gtttctatcc tcctgctcat    1140

cagctatctc tggacatgct gaaggtaact ctgatgtgtg aggtttttaga ctatggaaac    1200

taactctgtt cctgttgttt gcactgacct ggacttctct cccttactgc tagcgacttt    1260

caacagcaaa tgatgaaata gtagaagttc tcctttccaa acaccaagtg ttagctgcct    1320

taaggtttat ccggggcatt ggtggccatg acaacatttc tgcacgaaaa tttttagatg    1380

ctgcaaagca gactgaagac aacatgcttt tctatacaat attccgcttt tttgaacagc    1440

gaaaccagcg tttgcgaggg agccccaatt tcacaccagg ggaacactgt gaagaacatg    1500

ttgctttttt caaacagatt tttggagacc aagctctaat gaggcctaca acattctgaa    1560

atcacttgct gttttttat ataaaaatgt gtacaaagtt aatttattgc attaataaag    1620

ctctttaaac tataaaatgt taaaaagtg                                      1649
```

<210> 4

<211> 1861

<212> DNA

<213> Homo sapiens

<400> 4

```
gaaaaagttg tatctaagta gagaaagtgt tttccgctca tatcctgggt ccacatcgaa     60

gaattcagtc cttgtggatg aactgtaaac agcacccttc tctaagatg  ccgaagatca    120

tagtttgtgg ttttttttctt tcaggcggtg gaagcagggc agagccgaag cagcccgctc    180

ctcaagaggc cggtgcggac ccaggcggtg ctggaccagt cagatgtgta cacccatgtc    240

ctgtcagcct tcgtggaaaa gaaggtgggc cgcagctttc cgcctcttct ggactgagaa    300

tgctcaaaac aaggaagttg ctgaaaacga ggagacttca tgtgattaga gtcacttgaa    360

gtgattagaa tcactggagt ttccttgggt gaggccctag agctggtagt ttggcttcta    420

atgctgaggc ctaaagcata attgttgacg ggtggttctg gagcgatttg tgcaaaacca    480

gtgaaagatg aacactgggc cattttaaga tggaaacaag gtggggttg  gatagagagt    540
```

4

```
tatatgcagc ctctttttgca cctcgttggt atttgtaaga ccacatttt ttctccctag       600

gagatgcctc ataaatttgt gatagccgtg ctgatggaat acattcgttc tcttaaccag       660

tttcagattg cagtacagcc ttcaaatcat ctgggcccaa gttaaaacag aaggaattta       720

aaaaaaaaac acagtcactg tcttagaaga tgactcatat gctaagacag gtctgcctcc       780

ctgactcaga atgctgagtg actcctgaca ttattagttg gaatgggaag tgtaaggtca       840

agttggggtc tttacctgca tgacgaaacc acttcttgta atgacagact tttactgtgt       900

tggttagaat agccagtcct tggggagcct ctagtctgtt gtagctgaat gatttggaag       960

tgttctttca cttttttactt ttgtcctcag cattacctac atgaacttgt tatcaaaacc      1020

cttgtccagc acaacctctt ttatatgctg catcagttcc tgcagtacca cgtcctcagc      1080

gactccaaac ctttggcttg tctgctgtta tccctagaga gtttctatcc tcctgctcat      1140

cagctatctc tggacatgct gaaggtaact ctgatgtgtg aggttttaga ctatggaaac      1200

taactctgtt cctgttgttt gcactgacct ggacttctct cccttactgc tagcgacttt      1260

caacagcaaa tgatgaaata gtagaagttc tcctttccaa acaccaagtg ttagctgcct      1320

taaggtttat ccggggcatt ggtggccatg acaacatttc tgcacgaaaa tttttagatg      1380

ctgcaaagca gactgaagac aacatgcttt tctatacaat attccgcttt tttgaacagc      1440

gaaaccagcg tttgcgaggg agccccaatt tcacaccagg tgagaatgca atgaaaagac      1500

ttggggtaac catagcctca aagagtagca gagggcactg gcagctggtg ggcgaggacc      1560

ctgggttagc attttttgtaa acaacacaat ttgataacag cccacctagc ccttggccca      1620

ttatttgtag tagagtgaat tcagtatact gacagaatct ggattatgct ctggaactca      1680

ccgaggaggt gtgttttgag tcaagacaca tttaggaccc agatcaggca cagcccatct      1740

cttatagcag atcttggaat atctcttaaa gccaggaata agacggcaaa tggtggctaa      1800

gggttttaaa gggtctgggg cttattaagg tttcagtttt atgaagtata cattggttga      1860

t                                                                       1861
```

<210> 5

<211> 214

<212> DNA

<213> Homo sapiens

<400> 5
```
gtaagggaac tttggcgact tagtgcgatc actgggagaa ttgtagagtc cactggagag        60

aaagaaaaat ggtcaaaaag agcccagaga gttcctgggg gaaaacacac cgcagcccag       120

acctattcat aactgcacag ctggtacttc cagaggcaca tgcaccaggg gcacgtggtt       180

ctctttgctg acaagattta ttaaaagaaa agag                                   214
```

<210> 6

<211> 1934

5

<212> DNA

<213> Homo sapiens

<400> 6

```
aagtcaacga aaggttccgt tgtccttgac cacgtattcc atcacgtaaa ccttgtggag     60

atagattatt ttgggctacg ttactgtgac agaagccatc agacgtattg gctggatcct    120

gcaaaaaccc ttgctgaaca caaagaactg atcaacactg gacctccata tactttgtat    180

tttggtatta aattctatgc tgaagatcca tgtaaactta aagaagaaat aaccagatat    240

cagttttct tgcaggtgaa gcaagatgtc cttcagggcc gtctgccctg tcccgtcaac    300

actgctgctc agctgggagc gtatgccatc cagtcggagc ttggagatta tgacccatat    360

aaacatactg caggatatgt atctgagtac cggtttgttc ctgatcagaa ggaagaactt    420

gaagaagcca tagaaaggat tcataaaact ctaatgggtc agattccttc tgaggctgag    480

ctgaattact tgaggactgc caaatccctg gagatgtatg gcgttgacct ccatcccgtc    540

tatggagaaa acaagtctga gtatttctta ggattaactc cggttggtgt tgttgtgtac    600

aagaataaaa agcaagtggg gaagtatttc tggcctcgga ttacaaaggt tcacttcaag    660

gagactcaat ttgaactcag agtactggga aaagattgta acgaaacctc attctttttt    720

gaagctcgga gtaaaactgc ttgcaagcac ctctggaagt gcagtgtgga acatcataca    780

tttttttagaa tgccagaaaa tgaatccaat tcactgtcaa gaaaactcag caagtttgga    840

tccatacgtt ataagcaccg ctacagtggc aggacagctt tgcaaatgag ccgagatctt    900

tctattcagc ttccccggcc tgatcagaat gtgacaagaa gtcgaagcaa gacttaccct    960

aagcgaatag cacaaacaca gccagctgaa tcaaacacca tcagtaggat aactgcaaac   1020

atggaaaatg gagaaaatga aggaacaatt aaaattattg caccttcacc agtaaaaagc   1080

tttaagaaag caaagaatga aaatagccct gatacccaaa gaagcaaatc tcatgcaccg   1140

tgggaagaaa atggcccca gagtggactc tacaattctc ccagtgatcg cactaagtcg   1200

ccaaagttcc cttacacgcg tcgccgaaac ccctcctgtg gaagtgacaa tgattctgta   1260

cagcctgtga ggaggaggaa agcccataac agtggtgaag attcagatct taagcaaagg   1320

aggaggtcac gttcacgctg taacaccagc agtggtagtg aatcagaaaa ttctaataga   1380

gaacaccgga aaaagagaaa cagaatacgg caggagaatg atatggttga ttcagcgcct   1440

cagtgggaag ctgtattaag gagacaaaag gaaaaaaacc aagccgaccc caacagcagg   1500

cgatccagac acagatctcg ttcgagaagc cccgatatcc aagcaaaaga gagttatgg   1560

aagcacattc aaaaagaact tgtggatcca tccggattgt ccgaagaaca attaaaagag   1620

attccataca ctaaaataga gtgagtgcct ttcagaatct tctcaccaaa gctttattag   1680

tgcttgtgag taatccattc taattcttca attgtgttcc agacagtgct ttaatttgtc   1740

tttacatttt aaccaaaact aggtgacagt agcgaaagag gaagaaaagt gtgcattaaa   1800

gctacttatt ctacactata atcactatca tctcttatta gccacctctt tgtacttggt   1860

aggtacaagg gggctttcc tgattaatgt cagttttaaa ataaattctt ttctgagatt   1920
```

ctcactgaaa aaat   1934

<210> 7

<211> 2353

<212> DNA

<213> Homo sapiens

<400> 7
aagtcaacga aaggttccgt tgtccttgac cacgtattcc atcacgtaaa ccttgtggag   60

atagattatt ttgggctacg ttactgtgac agaagccatc agacgtattg gctggatcct   120

gcaaaaaccc ttgctgaaca caaagaactg atcaacactg gacctccata tactttgtat   180

tttggtatta aattctatgc tgaagatcca tgtaaactta agaagaaat aaccagatat   240

cagtttttct tgcaggtgaa gcaagatgtc cttcagggcc gtctgccctg tcccgtcaac   300

actgctgctc agctgggagc gtatgccatc cagtcggagc ttggagatta tgacccatat   360

aaacatactg caggatatgt atctgagtac cggtttgttc ctgatcagaa ggaagaactt   420

gaagaagcca tagaaaggat tcataaaact ctaatgggtc agattccttc tgaggctgag   480

ctgaattact tgaggactgc caaatccctg gagatgtatg gcgttgacct ccatcccgtc   540

tatggagaaa acaagtctga gtatttctta ggattaactc cggttggtgt tgttgtgtac   600

aagaataaaa agcaagtggg gaagtatttc tggcctcgga ttacaaaggt tcacttcaag   660

gagactcaat ttgaactcag agtactggga aaagattgta acgaaacctc attctttttt   720

gaagctcgga gtaaaactgc ttgcaagcac ctctggaagt gcagtgtgga acatcataca   780

ttttttagaa tgccagaaaa tgaatccaat tcactgtcaa gaaaactcag caagtttgga   840

tccatacgtt ataagcaccg ctacagtggc aggacagctt tgcaaatgag ccgagatctt   900

tctattcagc ttccccggcc tgatcagaat gtgacaagaa gtcgaagcaa gacttaccct   960

aagcgaatag cacaaacaca gccagctgaa tcaaacacca tcagtaggat aactgcaaac   1020

atggaaaatg gagaaaatga aggaacaatt aaaattattg caccttcacc agtaaaaagc   1080

tttaagaaag caaagaatga aaatagccct gatacccaaa gaagcaaatc tcatgcaccg   1140

tgggaagaaa atggcccccca gagtggactc tacaattctc ccagtgatcg cactaagtcg   1200

ccaaagttcc cttacacgcg tcgccgaaac ccctcctgtg gaagtgacaa tgattctgta   1260

cagcctgtga ggaggaggaa agcccataac agtggtgaag attcagatct taagcaaagg   1320

aggaggtcac gttcacgctg taacaccagc agtggtagtg aatcagaaaa ttctaataga   1380

gaacaccgga aaaagagaaa cagaatacgg caggagaatg atatggttga ttcagcgcct   1440

cagtgggaag ctgtattaag gagacaaaag gaaaaaaacc aagccgaccc caacagcagg   1500

cgatccagac acagatctcg ttcgagaagc cccgatatcc aagcaaaaga agagttatgg   1560

aagcacattc aaaagaact tgtggatcca tccggattgt ccgaagaaca attaaaagag   1620

attccataca ctaaaataga gacacaaggt gacccaatcc gcatcaggca ttctcattcg   1680

7

```
ccacgaagtt accgccagta tcgcaggtcc cagtgttcag atggggagcg atcagttctc      1740

tcggaagtga attcaaaaac agatcttgta ccaccacttc cggtgaccca ttcttcggat      1800

gctcagggtt ctggggatgc tacagttcat cagagaagaa atgggtctaa agatagcctg      1860

atggaagaaa aacctcagac atctacaaac aacctggctg gaaaacacac agcaaaaaca      1920

ataaaaacta tacaagcttc ccgcctcaag acagagactt gatcctgatg aagggtcaag      1980

ggtaggggtg ggaaggttgt gtgcgccact ggtacttttg aaactgtgaa ataggtatct      2040

taattcaaat ctcagacctg caagtatttc ttcagcatga gaaaatacat tatcttttgc      2100

ttctttttttt tttttttttg agatgttatc actctgtcgc ccaggctgga gtgcagcggc      2160

accgtgtcag ctcaccgcag cctccactta ctgggttaag cgattctcct gtctcaggct      2220

accgagcagc tgggattaca ggcgtgcacc acaacacccg gctaattctt tttgtatttt      2280

tagtagagac agggctttgc catgttggag gctggtctcg aactcctgac ctcaagtgat      2340

ccgcctgcct cag                                                        2353
```

<210> 8

<211> 2500

<212> DNA

<213> Homo sapiens

<400> 8
```
gacatgggct gtttctgcgc tgttccggaa gaattttact gcgaagtttt gctcctggat        60

gaatccaagt taacccttac cacccagcag cagggcatca agaagtcaac gaaaggttcc       120

gttgtccttg accacgtatt ccatcacgta aaccttgtgg agatagatta ttttgggcta       180

cgttactgtg acagaagcca tcagacgtat tggctggatc ctgcaaaaac ccttgctgaa       240

cacaaagaac tgatcaacac tggacctcca tatactttgt attttggtat taaattctat       300

gctgaagatc catgtaaact taaagaagaa ataaccagat atcagttttt cttgcaggtg       360

aagcaagatg tccttcaggg ccgtctgccc tgtcccgtca acactgctgc tcagctggga       420

gcgtatgcca tccagtcgga gcttggagat tatgacccat ataaacatac tgcaggtat        480

gtatctgagt accggtttgt tcctgatcag aaggaagaac ttgaagaagc catagaaagg       540

attcataaaa ctctaatggg tcagattcct tctgaggctg agctgaatta cttgaggact       600

gccaaatccc tggagatgta tggcgttgac ctccatcccg tctatggaga aaacaagtct       660

gagtatttct taggattaac tccggttggt gttgttgtgt acaagaataa aaagcaagtg       720

gggaagtatt tctggcctcg gattacaaag gttcacttca aggagactca atttgaactc       780

agagtactgg gaaaagattg taacgaaacc tcattctttt ttgaagctcg gagtaaaact       840

gcttgcaagc acctctggaa gtgcagtgtg gaacatcata cattttttag aatgccagaa       900

aatgaatcca attcactgtc aagaaaactc agcaagtttg gatccatacg ttataagcac       960

cgctacagtg gcaggacagc tttgcaaatg agccgagatc tttctattca gcttccccgg      1020

cctgatcaga atgtgacaag aagtcgaagc aagacttacc ctaagcgaat agcacaaaca      1080
```

8

```
cagccagctg aatcaaacac catcagtagg ataactgcaa acatggaaaa tggagaaaat   1140

gaaggaacaa ttaaaattat tgcaccttca ccagtaaaaa gctttaagaa agcaaagaat   1200

gaaaatagcc ctgataccca aagaagcaaa tctcatgcac cgtgggaaga aaatggcccc   1260

cagagtggac tctacaattc tcccagtgat cgcactaagt cgccaaagtt cccttacacg   1320

cgtcgccgaa acccctcctg tggaagtgac aatgattctg tacagcctgt gaggaggagg   1380

aaagcccata acagtggtga agattcagat cttaagcaaa ggaggaggtc acgttcacgc   1440

tgtaacacca gcagtggtag tgaatcagaa aattctaata gagaacaccg gaaaaagaga   1500

aacagaatac ggcaggagaa tgatatggtt gattcagcgc ctcagtggga agctgtatta   1560

aggagacaaa aggaaaaaaa ccaagccgac cccaacagca ggcgatccag acacagatct   1620

cgttcgagaa gccccgatat ccaagcaaaa gaagagttat ggaagcacat tcaaaaagaa   1680

cttgtggatc catccggatt gtccgaagaa caattaaaag agattccata cactaaaata   1740

gagtgagtgc ctttcagaat cttctcacca aagctttatt agtgcttgac acaaggtgac   1800

ccaatccgca tcaggcattc tcattcgcca cgaagttacc gccagtatcg caggtcccag   1860

tgttcagatg gggagcgatc agttctctcg gaagtgaatt caaaaacaga tcttgtacca   1920

ccacttccgg tgacccattc ttcggatgct cagggttctg gggatgctac agttcatcag   1980

agaagaaatg ggtctaaaga tagcctgatg gaagaaaaac ctcagacatc tacaaacaac   2040

ctggctggaa aacacacagc aaaaacaata aaaactatac aagcttcccg cctcaagaca   2100

gagacttgat cctgatgaag ggtcaagggt aggggtggga aggttgtgtg cgccactggt   2160

acttttgaaa ctgtgaaata ggtatcttaa ttcaaatctc agacctgcaa gtatttcttc   2220

agcatgagaa aatacattat cttttgcttc ttttttttttt tttttttgaga tgttatcact   2280

ctgtcgccca ggctggagtg cagcggcacc gtgtcagctc accgcagcct ccacttactg   2340

ggttaagcga ttctcctgtc tcaggctacc gagcagctgg gattacaggc gtgcaccaca   2400

acacccggct aattcttttt gtatttttag tagagacagg ctttgccat gttggaggct   2460

ggtctcgaac tcctgacctc aagtgatccg cctgcctcag                        2500
```

<210> 9

<211> 947

<212> DNA

<213> Homo sapiens

<400> 9
```
gaaagatgat actaggtcag gaaatagcat ttgaaagtca ttctcatctg gagggatgaa    60

gccaagataa ggcggaacca gggaaaagct ttaagaaagc aaagaatgaa aatagccctg   120

atacccaaag aagcaaatct catgcaccgt gggaagaaaa tggcccccag agtggactct   180

acaattctcc cagtgatcgc actaagtcgc caaagttccc ttacacgcgt cgccgaaacc   240

cctcctgtgg aagtgacaat gattctgtac agcctgtgag gaggaggaaa gcccattaac   300
```

9

```
agtggtgaag tttcagatct taaggcaaag ggagggaggt cacgttcacg ctgtaacacc    360

agcagtggta gtgaatcaga aaattctaat agagaacacc ggaaaaagag aaacagaata    420

cggcaggaga atgatatggt tgattcagcg cctcagtggg aagctgtatt aaggagacaa    480

aaggaaaaaa accaagccga ccccaacagc aggcgatcca gacacagatc tcgttcgaga    540

agccccgata tccaagcaaa agaagagtta tggaagcaca ttcaaaaaga acttgtggat    600

ccatccggat tgtccgaaga acaattaaaa gagattccat acactaaaat agagtgagtg    660

cctttcagaa tcttctcacc aaagctttat tagtgcttgt gagtaatcca ttctaattct    720

tcaattgtgt tccagacagt gctttaattt gtctttacat tttaaccaaa actaggtgac    780

agtagcgaaa gaggaagaaa agtgtgcatt aaagctactt attctacact ataatcacta    840

tcatctctta ttagccacct ctttgtactt ggtaggtaca aggggggcttt tcctgattaa    900

tgtcagtttt aaaataaatt cttttctgag attctcactg aaaaaat                  947
```

```
<210>  10

<211>  1366

<212>  DNA

<213>  Homo sapiens


<400>  10
gaaagatgat actaggtcag gaaatagcat ttgaaagtca ttctcatctg gagggatgaa     60

gccaagataa ggcggaacca gggaaaagct ttaagaaagc aaagaatgaa aatagccctg    120

atacccaaag aagcaaatct catgcaccgt gggaagaaaa tggcccccag agtggactct    180

acaattctcc cagtgatcgc actaagtcgc caaagttccc ttacacgcgt cggcgaaacc    240

cctcctgtgg aagtgacaat gattctgtac agcctgtgag gaggaggaaa gcccattaac    300

agtggtgaag tttcagatct taaggcaaag ggagggaggt cacgttcacg ctgtaacacc    360

agcagtggta gtgaatcaga aaattctaat agagaacacc ggaaaaagag aaacagaata    420

cggcaggaga atgatatggt tgattcagcg cctcagtggg aagctgtatt aaggagacaa    480

aaggaaaaaa accaagccga ccccaacagc aggcgatcca gacacagatc tcgttcgaga    540

agccccgata tccaagcaaa agaagagtta tggaagcaca ttcaaaaaga acttgtggat    600

ccatccggat tgtccgaaga acaattaaaa gagattccat acactaaaat agagacacaa    660

ggtgacccaa tccgcatcag gcattctcat tcgccacgaa gttaccgcca gtatcgcagg    720

tcccagtgtt cagatgggga gcgatcagtt ctctcggaag tgaattcaaa aacagatctt    780

gtaccaccac ttccggtgac ccattcttcg gatgctcagg gttctgggga tgctacagtt    840

catcagagaa gaaatgggtc taaagatagc ctgatggaag aaaaacctca gacatctaca    900

aacaacctgg ctggaaaaca cacagcaaaa acaataaaaa ctatacaagc ttcccgcctc    960

aagacagaga cttgatcctg atgaagggtc aagggtaggg gtgggaaggt tgtgtgcgcc   1020

actggtactt ttgaaactgt gaaataggta tcttaattca aatctcagac ctgcaagtat   1080

ttcttcagca tgagaaaata cattatcttt tgcttctttt ttttttttt ttgagatgtt   1140
```

10

```
atcactctgt cgcccaggct ggagtgcagc ggcaccgtgt cagctcaccg cagcctccac    1200

ttactgggtt aagcgattct cctgtctcag gctaccgagc agctgggatt acaggcgtgc    1260

accacaacac ccggctaatt cttttttgtat ttttagtaga gacagggctt tgccatgttg    1320

gaggctggtc tcgaactcct gacctcaagt gatccgcctg cctcag                   1366
```

<210> 11

<211> 422

<212> DNA

<213> Homo sapiens

<400> 11
```
aatcttcata atccccatgt gtcaaaggag agaccaggtg gaggtaactg aatcatgggg     60

gtggtttccc caggctgttt ttgtgatagt gagtgagttc tcatgagatc tgatggtttt    120

ataaggggct cttccctcct ttgcttgtga agaaggtgcc tttttttcccc tttgccttct   180

gccatgattg taagtttcct gaggcctccc cagccaagct gaactgtgag tcaattaaac    240

ctctttctt cgtaaattac ccagtcttga gcagttcttt acagcagtgt gaaaacagag     300

gaatacaccc atacatgcta ttctctgccc agaagccagg gggagcctgc cattaaaatg    360

aaagtcactc cttgactcag aaccctcaaa tagctttcat ctcacccaga aaaaaaagaa    420

aa                                                                   422
```

<210> 12

<211> 1532

<212> DNA

<213> Homo sapiens

<400> 12
```
aggtttctgc acaggaatat cgagagcgtc atgaacccga gctatagaga aaggagatga     60

ggcgtgagcc accgcacccg gctgacaagt gtccttctaa gaaacacaca gaggagaaga    120

cacagaagag gagagcacca tgtgatggta gacacagaaa ttggagttct acagccacaa    180

gccaaggaac tcctggagcc accaggagat ggaagatgca aagaactgat tttctctcag    240

agcctctgga gggagtgtgg ccctggtgac accttgattt tggacttctg gcctacagaa    300

ccatgcacac aggaggactt catttcccag gtctccttgc agtgaagttg aggccatgtg    360

actggtcttg ggccaatgga atgggtgcag aagggacaca gcccatttct agactcagcc    420

tgaaatgtcc tccataatcc ttactctttc tcccttcact cactggctgc aggaagctga    480

gaattatcct tggacttaca taaagcattt tggactttat gtaagtaaca acctgttgta    540

ttaagctact aagattttac ggttgtttgt taaatcagct aaccttaaac atcctaacaa    600

ctacaaatag aatacctgtt actgcataca taaaaataca aaaattagct ggatgtggtc    660
```

11

```
ccacctgtag tcccagctac tcgggaggct gaggcaggag aattgcttga acctgggagg      720

cggaggttgt ggtgagctga gatcgcacca ctgcactgca gcctgggcaa cagagcagga      780

ctctatctca aaaaaaaaac acaataaaca tttcttacct actgtagttt ttgtgggtca      840

ggaatctggg agcagcttag ttggatgatt tctgctcaca gtgttttatg aggttgcagt      900

caagatgttg gctggggctg tagtcatctg gagatttaac tacggctgga ggatccactt      960

caccatggtt cactcacctg gtgctggttg ctggcaggaa atttcagctc ttctcttata     1020

tggatctctt cacagattgc ttgagtgtcc tcaccgtatg gtgactggct tcctttacag     1080

aaatcagttg aagggaatgg gcaagtaaga aacagcaatg cttttatga cctagtcctg      1140

aagttcccca ccattactta tgttcattgg aagccagttg ctaaggagag cctgcactca     1200

aagattgggg aaatagactt tatctttcaa agtgttgaag aatttgcaga cgtattttaa     1260

aaccaccaca caatccatca acacatcatg tcggctctat tcttgaaata gatccagaat     1320

ttgaccactt ttcaccatct ccattgctat tacccagatc taatcaacac catcacttgc     1380

ctggactaga gatttcctcc tcactgggct ctctgcttct atctttagcc cattgctatg     1440

atttggctgt gtccccaccc aaaatctcat cttgaattat aatcttcata atccccatgt     1500

gtcaaaggag agaccaggtg gaggtaactg aa                                   1532
```

<210> 13

<211> 1753

<212> DNA

<213> Homo sapiens

<400> 13
```
tttcttaggg ttttttttttg agttggagcc tcgctctgtc ccccaggctg gagtgcagtg       60

atgtgatctc ggctcactgc aacctctgcc tcccaggttc aagtgattct cctgcctcag      120

cctccctagt agctgcgact acaggcatgt gccaccatgc ctggctaacg ttttgtattt      180

ttgagtagag acagggtttc accatgttgg ccaggctatt ctcgaactcc tgacctcaag      240

tgatccacct gcctcggctt cccaaagttt ctgggattac aggcgtgagc caccgcaccc      300

ggctgacaag tgtccttcta agaaacacac agaggagaag acacagaaga ggagagcacc      360

atgtgatggt agacacagaa attggagttc tacagccaca agccaggaa ctcctggagc      420

caccaggaga tggaagatgc aaagaactga ttttctctca gagcctctgg agggagtgtg      480

gccctggtga caccttgatt ttggacttct ggcctacaga accatgcaca caggaggact      540

tcatttccca ggtctccttg cagtgaagtt gaggccatgt gactggtctt gggccaatgg      600

aatgggtgca gaagggacac agcccatttc tagactcagc ctgaaatgtc ctccataatc      660

cttactcttt ctcccttcac tcactggctg caggaagctg agaattatcc ttggacttac      720

ataaagcatt ttggacttta tgtaagtaac aacctgttgt attaagctac taagatttta      780

cggttgtttg ttaaatcagc taaccttaaa catcctaaca actacaaata gaatacctgt      840

tactgcatac ataaaaatac aaaaattagc tggatgtggt cccacctgta gtcccagcta      900
```

12

```
ctcgggaggc tgaggcagga gaattgcttg aacctgggag gcggaggttg tggtgagctg     960

agatcgcacc actgcactgc agcctgggca acagagcagg actctatctc aaaaaaaaaa    1020

cacaataaac atttcttacc tactgtagtt tttgtgggtc aggaatctgg gagcagctta    1080

gttggatgat ttctgctcac agtgttttat gaggttgcag tcaagatgtt ggctggggct    1140

gtagtcatct ggagatttaa ctacggctgg aggatccact tcaccatggt tcactcacct    1200

ggtgctggtt gctggcagga aatttcagct cttctcttat atggatctct tcacagattg    1260

cttgagtgtc ctcaccgtat ggtgactggc ttcctttaca gaaatcagtt gaagggaatg    1320

ggcaagtaag aaacagcaat gctttttatg acctagtcct gaagttcccc accattactt    1380

atgttcattg gaagccagtt gctaaggaga gcctgcactc aaagattggg gaaatagact    1440

ttatctttca aagtgttgaa gaatttgcag acgtatttta aaaccaccac acaatccatc    1500

aacacatcat gtcggctcta ttcttgaaat agatccagaa tttgaccact tttcaccatc    1560

tccattgcta ttacccagat ctaatcaaca ccatcacttg cctggactag agatttcctc    1620

ctcactgggc tctctgcttc tatctttagc ccattgctat gatttggctg tgtccccacc    1680

caaaatctca tcttgaatta taatcttcat aatcccatg tgtcaaagga gagaccaggt    1740

ggaggtaact gaa                                                        1753
```

<210>   14
<211>   1832
<212>   DNA
<213>   Homo sapiens

<400>   14
```
gggttttgcg ggtataatta cattcaggat ctcaggatac tgcattatct gtgtgacccc      60

taaatctgat gacaagtgtc tgttttttgt ttttgttttt gagacagagc ctcgctctgt     120

cacccaggct ggagtgctgt ggtgtgatct cggctcactg caacctccgc ctcccaggtt     180

caagcaattc tctgcctcag cctcccgagt aaatgtgatt acaggcaggc gcctgccagc     240

acacccagct gattttagta tttttagtag agatgggggtt tcaccatctt ggccaggctg     300

gtcttgaatt cctgacctcg tgatccaccc acttcagctt cccaaagttc tgggattaca     360

ggcgtgagcc accgcacccg gctgacaagt gtccttctaa gaaacacaca gaggagaaga     420

cacagaagag gagagcacca tgtgatggta gacacagaaa ttggagttct acagccacaa     480

gccaaggaac tcctggagcc accaggagat ggaagatgca aagaactgat tttctctcag     540

agcctctgga gggagtgtgg ccctggtgac accttgattt tggacttctg gcctacagaa     600

ccatgcacac aggaggactt catttcccag gtctccttgc agtgaagttg aggccatgtg     660

actggtcttg ggccaatgga atgggtgcag aagggacaca gcccatttct agactcagcc     720

tgaaatgtcc tccataatcc ttactctttc tcccttcact cactggctgc aggaagctga     780

gaattatcct ggacttaca taaagcattt tggactttat gtaagtaaca acctgttgta     840
```

13

```
ttaagctact aagattttac ggttgtttgt taaatcagct aaccttaaac atcctaacaa    900

ctacaaatag aatacctgtt actgcataca taaaaataca aaaattagct ggatgtggtc    960

ccacctgtag tcccagctac tcgggaggct gaggcaggag aattgcttga acctgggagg   1020

cggaggttgt ggtgagctga gatcgcacca ctgcactgca gcctgggcaa cagagcagga   1080

ctctatctca aaaaaaaac acaataaaca tttcttacct actgtagttt ttgtgggtca   1140

ggaatctggg agcagcttag ttggatgatt tctgctcaca gtgttttatg aggttgcagt   1200

caagatgttg gctggggctg tagtcatctg gagatttaac tacggctgga ggatccactt   1260

caccatggtt cactcacctg gtgctggttg ctggcaggaa atttcagctc ttctcttata   1320

tggatctctt cacagattgc ttgagtgtcc tcaccgtatg gtgactggct ccttttacag   1380

aaatcagttg aagggaatgg gcaagtaaga aacagcaatg ctttttatga cctagtcctg   1440

aagttcccca ccattactta tgttcattgg aagccagttg ctaaggagag cctgcactca   1500

aagattgggg aaatagactt tatctttcaa agtgttgaag aatttgcaga cgtattttaa   1560

aaccaccaca caatccatca acacatcatg tcggctctat tcttgaaata gatccagaat   1620

ttgaccactt ttcaccatct ccattgctat tacccagatc taatcaacac catcacttgc   1680

ctggactaga gatttcctcc tcactgggct ctctgcttct atctttagcc cattgctatg   1740

atttggctgt gtccccaccc aaaatctcat cttgaattat aatcttcata atccccatgt   1800

gtcaaaggag agaccaggtg gaggtaactg aa                                 1832
```

<210> 15

<211> 10394

<212> DNA

<213> Homo sapiens

<400> 15

```
cgttgtttgg cgtgtttttt ttttttgtttt ttgtcactgc ctgcctgggt cctgcccgag     60

gtctccatcc tcggtttccc tgtccttgcc ccgggccctg ggagtgctct ggaaggctgc    120

gcagtattgg aggggacaga atgaccttcc ggccttgagt ccctggggag cagatggacc    180

ctactggaag tcagttggat tcagatttct ctcagcaaga tactccttgc ctgataattg    240

aagattctca gcctgaaagc caggttctag aggatgattc tggttctcac ttcagtatgc    300

tatctcgaca ccttcctaat ctccagacgc acaaagaaaa tcctgtgttg gatgttgtgt    360

ccaatcctga acaaacagct ggagaagaac gaggagacgg taatagtggg ttcaatgaac    420

atttgaaaga aaacaaggtt gcagaccctg tggattcttc taacttggac acatgtggtt    480

ccatcagtca ggtcattgag cagttacctc agccaaacag gacaagcagt gttctgggaa    540

tgtcagtgga atctgctcct gctgtggagg aagagaaggg agaagagttg gaacagaagg    600

agaaagagaa ggaagaagat acttcaggca atactacaca ttcccttggt gctgaagata    660

ctgcctcatc acagttgggt tttggggttc tggaactctc ccagagccag gatgttgagg    720

aaaatactgt gccatatgaa gtggacaaag agcagctaca atcagtaacc accaactctg    780
```

14

```
gttataccag gctgtctgat gtggatgcta atactgcaat taagcatgaa gaacagtcca    840

acgaagatat ccccatagca gaacagtcca gcaaggacat ccctgtgaca gcacagccca    900

gtaaggatgt acatgttgta aaagagcaaa atccaccacc tgcaaggtca gaggacatgc    960

cttttagccc caaagcatct gttgctgcta tggaagcaaa agaacagttg tctgcacaag   1020

aacttatgga aagtggactg cagattcaga agtcaccaga gcctgaggtt ttgtcaactc   1080

aggaagactt gtttgaccag agcaataaaa cagtatcttc tgatggttgc tctactcctt   1140

caagggagga aggtgggtgt tctttggctt ccactcctgc caccactctg catctcctgc   1200

agctctctgg tcagaggtcc cttgttcagg acagtctttc cacgaattct tcagatcttg   1260

ttgctccttc tcctgatgct ttccgatcta ctccttttat cgttcctagc agtcccacag   1320

agcaagaagg gagacaagat aagccaatgg acacgtcagt gttatctgaa gaaggaggag   1380

agccttttca gaagaaactt caaagtggtg aaccagtgga gttagaaaac ccccctctcc   1440

tgcctgagtc cactgtatca ccacaagcct caacaccaat atctcagagc acaccagtct   1500

tccctcctgg gtcacttcct atcccatccc agcctcagtt ttctcatgac atttttattc   1560

cttccccaag tctggaagaa caatcaaatg atgggaagaa agatggagat atgcatagtt   1620

catctttgac agttgagtgt tctaaaactt cagagattga accaaagaat tcccctgagg   1680

atcttgggct atctttgaca ggggattctt gcaagttgat gctttctaca agtgaatata   1740

gtcagtcccc aaagatggag agcttgagtt ctcacagaat tgatgaagat ggagaaaaca   1800

cacagattga ggatacggaa cccatgtctc cagttctcaa ttctaaattt gttcctgctg   1860

aaaatgatag tatcctgatg aatccagcac aggatggtga agtacaactg agtcagaatg   1920

atgacaaaac aaagggagat gatacagaca ccagggatga cattagtatt ttagccactg   1980

gttgcaaggg cagagaagaa acggtagcag aagatgtttg tattgatctc acttgtgatt   2040

cggggagtca ggcagttccg tcaccagcta ctcgatctga ggcactttct agtgtgttag   2100

atcaggagga agctatggaa attaaagaac accatccaga ggaggggtct tcagggtctg   2160

aggtggaaga aatccctgag acaccttgtg aaagtcaagg agaggaactc aaagaagaaa   2220

atatggagag tgttccgttg cacctttctc tgactgaaac tcagtcccaa gggttgtgtc   2280

ttcaaaagga aatgccaaaa aaagaatgct cagaagctat ggaagttgaa accagtgtga   2340

ttagtattga ttcccctcaa aagttggcaa tacttgacca agaattggaa cataaggaac   2400

aggaagcttg ggaagaagct acttcagagg actccagtgt tgtcattgta gatgtgaaag   2460

agccatctcc cagagttgat gtttcttgtg aacctttgga gggagtggag aagtgctcag   2520

attcccagtc atgggaggat attgctccag aaatagaacc atgtgctgag aatagattag   2580

acaccaagga agaaagagt gtagaatatg aaggagatct gaaatcaggg actgcagaaa   2640

cagaacctgt agagcaagat tcttcacagc cttccttacc tttagtgaga gcagatgatc   2700

ctttaagact tgaccaggag ttgcagcagc cccaaactca ggagaaaaca agtaattcat   2760

taacagaaga ctcaaaaatg gctaatgcaa agcagctaag ctcagatgca gaggcccaga   2820

agctggggaa gccctctgcc catgcctcac aaagcttctg tgaaagttct agtgaaaccc   2880
```

15

```
catttcattt cactttgcct aaagaaggtg atatcatccc accattgact ggtgcaaccc   2940

cacctcttat tgggcaccta aaattggagc ccaagagaca cagtactcct attggtatta   3000

gcaactatcc agaaagcacc atagcaacca gtgatgtcat gtctgaaagc atggtggaga   3060

cccatgatcc catacttggg agtggaaaag gggattctgg ggctgcccca gacgtggatg   3120

ataaattatg tctaagaatg aaactggtta gtcctgagac tgaggcgagt gaagagtctt   3180

tgcagttcaa cctggaaaag cctgcaactg gtgaaagaaa aaatggatct actgctgttg   3240

ctgagtctgt tgccagtccc cagaagacca tgtctgtgtt gagctgtatc tgtgaagcca   3300

ggcaagagaa tgaggctcga agtgaggatc cccccaccac acccatcagg gggaacttgc   3360

tccactttcc aagttctcaa ggagaagagg agaaagaaaa attggagggt gaccatacaa   3420

tcaggcagag tcaacagcct atgaagccca ttagtcctgt caaggaccct gtttctcctg   3480

cttcccagaa gatggtcata caagggccat ccagtcctca aggagaggca atggtgacag   3540

atgtgctaga agaccagaaa gaaggacgga gtactaataa ggaaaatcct agtaaggcct   3600

tgattgaaag gcccagccaa aataacatag gaatccaaac catggagtgt tccttgaggg   3660

tcccagaaac tgtttcagca gcaacccaga ctataaagaa tgtgtgtgag caggggacca   3720

gtacagtgga ccagaacttt ggaaagcaag atgccacagt tcagactgag aggggggagtg   3780

gtgagaaacc agtcagtgct cctggggatg atacagagtc gctccatagc cagggagaag   3840

aagagtttga tatgcctcag cctccacatg gccatgtctt acatcgtcac atgagaacaa   3900

tccgggaagt acgcacactt gtcactcgtg tcattacaga tgtgtattat gtggatggaa   3960

cagaagtaga aagaaaagta actgaggaga ctgaagagcc aattgtagag tgtcaggagt   4020

gtgaaactga agtttccct tcacagactg ggggctcctc aggtgacctg ggggatatca   4080

gctccttctc ctccaaggca tccagcttac accgcacatc aagtgggaca agtctctcag   4140

ctatgcacag cagtggaagc tcagggaaag gagccggacc actcagaggg aaaaccagcg   4200

ggacagaacc cgcagatttt gccttaccca gctcccgagg aggcccagga aaactgagtc   4260

ctagaaaagg ggtcagtcag acagggacgc cagtgtgtga ggaggatggt gatgcaggcc   4320

ttggcatcag acagggaggg aaggctccag tcacgcctcg tgggcgtggg cgaaggggcc   4380

gcccaccttc tcggaccact ggaaccagag aaacagctgt gcctggcccc ttgggcatag   4440

aggacatttc acctaacttg tcaccagatg ataaatcctt cagccgtgtc gtgccccgag   4500

tgccagactc caccagacga acagatgtgg gtgctggtgc tttgcgtcgt agtgactctc   4560

cagaaattcc tttccaggct gctgctggcc cttctgatgg cttagatgcc tcctctccag   4620

gaaatagctt tgtagggctc cgtgttgtag ccaagtggtc atccaatggc tacttttact   4680

ctgggaaaat cacacgagat gtcggagctg ggaagtataa attgctcttt gatgatgggt   4740

acgaatgtga tgtgttgggc aaagacattc tgttatgtga ccccatcccg ctggacactg   4800

aagtgacggc cctctcggag gatgagtatt tcagtgcagg agtggtgaaa ggacatagga   4860

aggagtctgg ggaactgtac tacagcattg aaaaagaagg ccaaagaaag tggtataagc   4920

gaatggctgt catcctgtcc ttggagcaag gaaacagact gagagagcag tatgggcttg   4980

gcccctatga agcagtaaca cctcttacaa aggcagcaga tatcagctta gacaatttgg   5040
```

16

```
tggaagggaa gcggaaacgg cgcagtaacg tcagctcccc agccacccct actgcctcca   5100

gtagcagcag cacaacccct acccgaaaga tcacagaaag tcctcgtgcc tccatgggag   5160

ttctctcagg caaaagaaaa cttatcactt ctgaagagga acggtcccct gccaagcgag   5220

gtcgcaagtc tgccacagta aaacctggtg cagtaggggc aggagagttt gtgagcccct   5280

gtgagagtgg agacaacacc ggtgaaccct ctgccctgga agagcagaga gggcctttgc   5340

ctctcaacaa gaccttgttt ctgggctacg catttctcct taccatggcc acaaccagtg   5400

acaagttggc cagccgctcc aaactgccag atggtcctac aggaagcagt gaagaagagg   5460

aggaattttt ggaaattcct cctttcaaca agcagtatac agaatcccag cttcgagcag   5520

gagctggcta tatccttgaa gatttcaatg aagcccagtg taacacagct taccagtgtc   5580

ttctaattgc ggatcagcat tgtcgaaccc ggaagtactt cctgtgcctt gccagtggga   5640

ttccttgtgt gtctcatgtc tgggtccatg atagttgcca tgccaaccag ctccagaact   5700

accgtaatta tctgttgcca gctgggtaca gccttgagga gcaaagaatt ctggactggc   5760

aaccccgtga aaatcctttc cagaatctga aggtactctt ggtatcagac caacagcaga   5820

acttcctgga gctctggtct gagatcctca tgactggtgg tgcagcctct gtgaagcagc   5880

accattcaag tgcccataac aaagatattg ctttaggggt atttgatgtg gtggtgacgg   5940

accccctcatg cccagcctcg gtgctgaagt gtgctgaagc attgcagctg cctgtggtgt   6000

cacaagagtg ggtgatccag tgcctcattg ttggggagag aattggattc aagcagcatc   6060

caaaatataa acacgattat gtttctcact aaagatactt ggtcttactg gtttttattcc   6120

ctgctatcgt ggagattgtg ttttaaccag gttttaaatg tgtcttgtgt gtaactggat   6180

tccttgcatg gatcttgtat atagttttat ttgctgaact tttatgataa aataaatgtt   6240

gaatctcttt ggttgtagta actgggattt cttcatctgt ttttttgagc ttaatctcag   6300

aacaaatgac aagacatagt actttctctg agtctttcaa caggcttatt cacttacgga   6360

ggacagctca ccaaggaaat tgaaaagtta agagtgaact ttattctgtg gcatcattcc   6420

caaaaggtta ttccagggtg tctaaaatgc tatgcttgca gaaactcagt ttaaggtagg   6480

tgaaggccca gattaacagt tgtgccaaaa gttgagtgga attgggcaca gctctgtttc   6540

ctgacagtta aaaaagacct catgctctct ctctgagctg agatcacagc tcacctgtgg   6600

gtactcccca actcttagag ctaaagggag aacgaaagga ccaactgcca tgaagggaca   6660

gtgaccataa gcttgatgga atgaccttcc gtaagataaa catgggaagc acaagtgaga   6720

acacctggaa atgttacacg ttctagtcaa agacccaata ttattattat tattattgtc   6780

acaatagctg gaagcagttc cttcccttcc tctggcatca ctgatccctg catggcttct   6840

cattctctaa agcaggggtc aacaaggttt ttttctgtaa agggtcaaag agtaaatatt   6900

tcaggctttg tgggccattt gatccatcac aactactcgc ctttgctgtg agggcatgaa   6960

agcaaccata gacaatgagt aaacaaatgg gcacggctgt gtttcagtaa aactgtacaa   7020

aaacagacag caggccatag tttgccagct cctgctccag agacagcagt ggaaagggtg   7080

atctttagtt gataatagca gggaataagt tgtcagagct tcccagtgtg tgtagaatat   7140
```

17

```
gtagtgatga aaaccagatg cagtgactat aacctgatgc cagaacactg cattcttttt      7200

cagtttggag ggcgttgttc agtgaatatt tcttttact tacactgata tgaatattga        7260

ttaccagtga tggctgggcc atattaagat aacttcaacc cctatggttt gtgtaagatg       7320

ggtaattggg cctgcaatct tcagtattta aaaatctaac aacttgatct caatttttc        7380

ttaaggacct ttttcttgga gaataatact tttttttttt ttttttttttt tgagacggaa     7440

tttcgctctt gttgcccagg ctggaatgca atggcacaat ctcagctcac tgcagcgtct      7500

gcttcccagg ttcaagcaat tctcctgtct cagcctcctg agtagctggg attacaggca       7560

catgccacca cacctggcta attttttgtat ttttagtaga atcgaggttt catcatgttg      7620

gtcaggctgg tctcaaactc ctgacttcag gtgatccgcc cgcctcggcc tcccaaagtg       7680

ctgggattac aggtgtgagc caccatgccc ggcctaagaa atacttttaa gtatatttc        7740

attagctaga attgcccaat ctgtgtaggt ataaattact tggtataggg agagagaaag       7800

cctatcttac ctgttgcttt cttacttggt ggtaacatcc agcagttagt ctatttataa       7860

acataattac ttttttcacat atgaaccata aaatatttaa ctttctgctc tatattgttt      7920

gtttaccgct gtatctccca cagcttgaac agtaccaagg tacgtagtag gtgctcaata       7980

aatgactatt gaataaatga acatatccaa caaatgttct caatgtaaag gatcagagat       8040

gccacatgtt ctccttgatg ggagagaccc ttccacatgg gaatgatggg aaggagttgt       8100

actcctggat gttcagtaac tgcttctagg agaaaaggta gagtcctatc actaagccgc       8160

agatatttat ttgtgtgtgg ctagaatggg atgttttgaa tcttctgtta caaccttggg       8220

aacgtggctg ttatttcaat ttatgagcca gaaattttca catcccgaaa ctacaaaaga      8280

gaaaaagagc cttattaagt gtcatgcttt cccaagacta ccttcaaaga aatatgaatc       8340

aggataacct gtgatctaaa taatgtcatc ttaaaactga agagtttctt ttgactcttc       8400

tgctacaata gcttagaaaa aaatctgctt gcagacattt tagagagaaa ggacaatgaa       8460

gtgattttct gaatgggaat gacagacctc tgggaagcca gctaccactg aatctcggta      8520

tcagtttttt ttaaagttta gagttagaag gggtggtcgc ctcctttcac agatgcggaa      8580

gctagggacc agcaaggcgg ggtgcccacg gctgcacagc tagttcatat cagaattggg       8640

agtggaaggc ccactgcctc ccagcatagc aatacataac ctagcaaagg acttaacacc       8700

tatctcactg tcaggttttt tagtatttta tgatgatgat gacttctact agaaaataac        8760

ctccattaaa attattaaag atggtcacac ctctatctct aagccttact tataaaatga       8820

gggtatttgg actaaagtct tcttccagtt ctagaattct acaactcatt aaaaagccac       8880

cttaaaaagt ctactgagtt acccaagggt tgctcctacc tgcccagagt tccaccagcc       8940

tgggtatagt atttgttata atctagtcgt aacagtagtt gagccaaatc tgagttgatc       9000

tgatgattcc gaacactgga gagaatcttg aacaggagtg aagactggcg gctaaagccc       9060

tgcagagaga aggactcagc tgtcattcca cttcagctca ccaactctcc atatggagga       9120

tgggggcgga gggaggagtt tcttggaaaa gccttgttca aaattctaca gaaccacctg       9180

gccttcccac attcctattt ctattagttc ctaaaatgac ttgtaccaaa tccatacatg       9240

catgacttcc tatgaaagta ctctttcatc agtaggaatt tagtagctgg tttccagtta       9300
```

```
atgtattttg tcaagtactg gggttgggga gaacccgttt tgattacaag cagataatta    9360

tctcagtgag atgggggtta gttcaaggaa gtaaggaggg gggaggatgt gaggaagtta    9420

gaacaaccca atgcttattt gatgggctga ataaactatt caggactgaa ctatttttga    9480

gcactgtgag gtggcacagt aattacctgc ttcaaaatca actgatacca acatttttat    9540

ctttgtatct tatctctgta cgtgtgtgta ttgaggaaat gctttactga ctcagaggaa    9600

agatcatgaa ttctccattg gcaaaaccac ctctgtcctt tcggcaaggc tgcatacttc    9660

caggcagacg caccttcacg agaatgctca gctgggcggc tccacgctca tccagtgggc    9720

ctaggttctg actgaccagc gaacaaaaac tgtgacagag atctaggatt tcattcaggc    9780

agtgaaacac ctacccggga aacagagttg gcattaggaa aggaaggaag gtacatccat    9840

gaagttaaag tgttaggaga acagtctgat taatagctga tctaattaat agctgacctc    9900

ccaaatctga caggatagac actgccacgt gcaaggcctg ccagcccctc agacgcacaa    9960

aatgcgtaaa acaaatgcat cctttcctgg ctaagcgagt attactctct tagccctgca    10020

ccaaacctcc aatctagcca catttaactc ttcatttctt agacccgcag agtgtcttcc    10080

tgcctctgag ctgtgagtgt tgttcccttt gcccgggatg ctcttgtttt taataccagt    10140

tcaagtccca ctctctcagt gaagcactcc cttccccact atagcctttta gtgaaccctc    10200

gtttcttgct tctttattat ctgtactgtt gtccacttgg caattgttca ggcctctgtg    10260

ttgttactga ttttttgtatg tatatatata tatatgtctt gtttttccaa ctagattgtg    10320

agctccttaa gggcagagcc atgaattata cctctttgta tccccagtgc cttgcataca    10380

gtaagcactc aata                                                      10394
```

```
<210>    16

<211>    6837

<212>    DNA

<213>    Homo sapiens


<400>    16
agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg      60

gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgaccccttc ccagcttccc     120

gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag     180

gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg     240

aggagggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct     300

cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca     360

ggtatcgcag gacttcccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg     420

ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt     480

gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct     540

gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga     600
```

19

```
tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct   660

agaccagttc cagcttcttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca   720

aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg gggggttgcc   780

tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catggggtca tgtataaaca   840

gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa   900

acaggggcca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg   960

cattggggga ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga   1020

agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc   1080

ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt   1140

tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac   1200

ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga   1260

acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga   1320

agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg   1380

agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc   1440

agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga   1500

tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa   1560

agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc   1620

tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct   1680

cttcacccca gctgtcctgg aaaagtacaa tgttgttttt aagtacttac tgagtgtgcg   1740

ccgggtgcaa gctgagctgc agcactgctg ggccctacaa atgcagcgca agcacctcaa   1800

gtcgaaccag actgatgcaa tcaagtggcg cctaagaaat cacatggcat ttttggtgga   1860

taatcttcag tactatctcc aggtagatgt gttggagtct cagttctccc agctgcttca   1920

tcagatcaat tctacccgag actttgaaag catccgattg gctcatgacc acttcctgag   1980

caatttgctg gctcaatcct ttatcctatt gaaacctgtg tttcactgcc tgaatgaaat   2040

cctagatctc tgtcacagtt tttgtttgct ggtcagtcag aacctaggcc cactggatga   2100

gcgtggagcc gcccagctga gcattctcgt gaagggcttt agccgccagt cttcactcct   2160

gttcaagatt ctctccagtg ttcggaatca tcagatcaac tcagatttgg ctcaactact   2220

gttacgacta gattataaca aatactatac ccaggctggt ggaactctgg gcagtttcgg   2280

gatgtgaaaa tttctggctc ataaattgaa ataacagcca cgttcccaag gttgtaacag   2340

aagattcaaa acatcccatt ctagccacac acaaataaat atctgcggct tagtgatagg   2400

actctacctt ttctcctaga agcagttact gaacatccag gagtacaact ccttcccatc   2460

attcccatgt ggaagggtct ctcccatcaa ggagaacatg tggcatctct gatcctttac   2520

attgagaaca tttgttggat atgttcattt attcaatagt catttattga gcacctacta   2580

cgtaccttgg tactgttcaa gctgtgggag atacagcggt agacaaacaa tatagagcag   2640

aaagttaaat attttatggt tcatatgtga aaaagtaatt atgtttataa atagactaac   2700

tgctggatgt taccaccaag taagaaagca acaggtaaga taggctttct ctctccctat   2760
```

```
accaagtaat ttatacctac acagattggg caattctagc taatgaaaat atacttaaaa   2820

gtatttctta ggccgggcat ggtggctcac acctgtaatc ccagcacttt gggaggccga   2880

ggcgggcgga tcacctgaag tcaggagttt gagaccagcc tgaccaacat gatgaaacct   2940

cgattctact aaaaatacaa aaattagcca ggtgtggtgg catgtgcctg taatcccagc   3000

tactcaggag gctgagacag gagaattgct tgaacctggg aagcagacgc tgcagtgagc   3060

tgagattgtg ccattgcatt ccagcctggg caacaagagc gaaattccgt ctcaaaaaaa   3120

aaaaaaaaaa aaaaagtatt attctccaag aaaaaggtcc ttaagaaaaa attgagatca   3180

agttgttaga ttttttaaata ctgaagattg caggcccaat tacccatctt acacaaacca   3240

tagggggttga agttatctta atatggccca gccatcactg gtaatcaata ttcatatcag   3300

tgtaagtaaa aagaaatatt cactgaacaa cgccctccaa actgaaaaag aatgcagtgt   3360

tctggcatca ggttatagtc actgcatctg gttttcatca ctacatattc tacacacact   3420

gggaagctct gacaacttat tccctgctat tatcaactaa agatcaccct ttccactgct   3480

gtctctggag caggagctgg caaactatgg cctgctgtct gtttttgtac agttttactg   3540

aaacacagcc gtgcccattt gtttactcat tgtctatggt tgctttcatg ccctcacagc   3600

aaaggcgagt agttgtgatg gatcaaatgg cccacaaagc ctgaaatatt tactctttga   3660

ccctttacag aaaaaaacct tgttgacccc tgctttagag aatgagaagc catgcaggga   3720

tcagtgatgc cagaggaagg gaaggaactg cttccagcta ttgtgacaat aataataata   3780

ataatattgg gtctttgact agaacgtgta acatttccag gtgttctcac ttgtgcttcc   3840

catgtttatc ttacggaagg tcattccatc aagcttatgg tcactgtccc ttcatggcag   3900

ttggtccttt cgttctccct ttagctctaa gagttgggga gtacccacag gtgagctgtg   3960

atctcagctc agagagagag catgaggtct tttttaactg tcaggaaaca gagctgtgcc   4020

caattccact caacttttgg cacaactgtt aatctgggcc ttcacctacc ttaaactgag   4080

tttctgcaag catagcattt tagacaccct ggaataacct tttgggaatg atgccacaga   4140

ataaagttca ctcttaactt ttcaatttcc ttggtgagct gtcctccgta agtgaataag   4200

cctgttgaaa gactcagaga aagtactatg tcttgtcatt tgttctgaga ttaagctcaa   4260

aaaaacagat gaagaaatcc cagttactac aaccaaagag attcaacatt tattttatca   4320

taaaagttca gcaaataaaa ctatatacaa gatccatgca aggaatccag ttacacacaa   4380

gacacattta aaacctggtt aaaacacaat ctccacgata gcaggaata aaaaccagtaa   4440

gaccaagtat ctttagtgag aaacataatc gtgtttatat tttggatgct gcttgaatcc   4500

aattctctcc ccaacaatga ggcactggat cacccactct tgtgacacca caggcagctg   4560

caatgcttca gcacacttca gcaccgaggc tgggcatgag gggtccgtca ccaccacatc   4620

aaatacccct aaagcaatat ctgcaaggag caagggaaag tgaagaagga aaggacactc   4680

aacttagccc tccattagaa agagagattt gattctaacc aatacatccc actctgcaca   4740

aaccaaagcc ctattatgtc aaacacactg ctactgatca tgaccaaagg cagagttata   4800

atcactatgt gctgaccttg tagaaatatt taacaaatat acgtccagtg cttcacttat   4860
```

21

```
gttgactcac ctcttgaagg tggtactttt cttctctaag aaacatggat acggtcaacc   4920

tattaggcct gagccttgga ccacaaggcc taacacctac aggtctaagg agatccctgg   4980

aacaaagaca ctacacacac tctttcaggt acctttgtta tgggcacttg aatggtgctg   5040

cttcacagag gctgcaccac cagtcatgag gatctcagac cagagctcca ggaagttctg   5100

ctgttggtct gataccaaga gtaccttcag attctggaaa ggattttcac ggggttgcct   5160

atgaaggaga caggaaagga ccttagcatg acaagtaata tccaacaaac tgcctttctg   5220

caaagggact catgtacatc tgaatgcttt caaaaataaa tgccccatca gacatagtgt   5280

ctcaagcctg taatcccagc actttgggag gctgtcgtgg ttggatctct tgggcctggg   5340

agttcgagac cagcctgggc aatgtggtga gaccccatct ctacaaaaga caacaaaaaa   5400

attagctggg tgtggtggcg agtgcctgta gtcccagcag cttgggaggc tgaggtaggg   5460

ggatcacttc agcctgggag gttgaggctg cagtaagtcg tcactgcgcc actgtactcc   5520

agcctaggtg acagagcaag acttcatctt aaaaaactaa gccctatatt agggtccccc   5580

ttctcttcct tctttctatg aatgatctgt attccttgca ttcctggctt tctaatttcc   5640

atgtttgttc tggggctgag aataatccaa atcatgctcc tgagcctata tatttttaat   5700

gcttgcttaa aacttagttc tctgacttta caggttgaga atattgaacc tatatacaaa   5760

tcttcacaca tttgcaaaag gttcctagcc aatgtaacct agggaaataa actagataaa   5820

ctcctgaagt catttcaaac ccactcaaat ttatcccaca gacattccaa tttctagaaa   5880

gctttactct ctcacctaga ttctcttccc tccaaagctt gctgtcctcc tgcctataca   5940

attctggatg ggcttcaaat acttaccagt ccagaattct ttgctcctca aggctgtacc   6000

cagctggcaa cagataatta cggtagttct ggagctggtt ggcatggcaa ctatcatgga   6060

cccagacatg agacacacaa ggaatcccac tggcaaggca caggaagtac ttccgggttc   6120

gacaatgctg atccgcaatt agaagacact ggtaagctgt gttacactgc aagaaaagaa   6180

gcagagccaa tgggtttggt gacttctgtg gaaagctcct aagcagcagc cataatgagc   6240

catgaagagc agatctgaag actcccaact actacccaaa atgtgattta gtctatcctg   6300

cccaaggcca ctcttctcac tggaaggccc aagtaatttc catagatgtt ctctctgcct   6360

cacctgcagc atactgagga cctaaatcct caacggacaa ccaaaaccta tgaactcagc   6420

ctttcaggct aaaaatcagc aaccctaata ggggtttcta ctactaaaca taaacatcaa   6480

tcttcttttg tcccagcaac agaaccatag ccattaacta acccaaggtc ctaccttctc   6540

ttccctatac acaacaaaaa ttctatttca tgcaaaaaca ttttggcagt ttctcagttc   6600

ctgaaatctc tggctacttt atccaggttc cccaacccct cccaggcctc ttctcaacac   6660

agcaagttgg ctcttatcat tgccactata ttaggttaca caaagaaact cctcacctgg   6720

gcttcattga aatcttcaag gatatagcca gctcctgctc gaagctggga ttctgtatac   6780

tgcttgttga aaggaggaat ttccaaaaat tctatattaa aaaaaaaaac caagata      6837
```

<210> 17

<211> 733

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 17
cacaatctcc acgatagcag ggaataaaac cagtaagacc aagtatcttt agtgagaaac   60

ataatcgtgt ttatattttg gatgctgctt gaatccaatt ctctccccaa caatgaggca   120

ctggatcacc cactcttgtg acaccacagg cagctgcaat gcttcagcac acttcagcac   180

cgaggctggg catgaggggt ccgtcaccac cacatcaaat acccctaaag caatatctgc   240

aaggagcaag ggaaagtgaa gaaggaaagg acactcaact tagccctcca ttagaaagag   300

agatttgatt ctaaccaata catcccactc tgcacaaacc aaagccctat tatgtcaaac   360

acactgctac tgatcatgac caaaggcaga gttataatca ctatgtgctg accttgtaga   420

aatatttaac aaatatacgt ccagtgcttc acttatgttg actcacctct tgaaggtggt   480

acttttcttc tctaagaaac atggatacgg tcaacctatt aggcctgagc cttggaccac   540

aaggcctaac acctacaggt ctaaggagat ccctggaaca aagacactac acacactctt   600

tcaggtacct ttgttatggg cacttgaatg gtgctgcttc acagaggctg caccaccagt   660

catgaggatc tcagaccaga gctccaggaa gttctgctgt tggtctgata ccaagagtac   720

cttcagattc tgg   733

<210> 18

<211> 734

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 18
gctagaattg cccaatctgt gtaggtataa attacttggt atagggagag agaaagccta   60

tcttacctgt tgctttctta cttggtggta acatccagca gttagtctat ttataaacat   120

aattactttt tcacatatga accataaaat atttaacttt ctgctctata ttgtttgtct   180

accgctgtat ctcccacagc ttgaacagta ccaaggtacg tagtaggtgc tcaataaatg   240

actattgaat aaatgaacat atccaacaaa tgttctcaat gtaaaggatc agagatgcca   300

catgttctcc ttgatgggag agacccttcc acatgggaat gatgggaagg agttgtactc   360

ctggatgttc agtaactgct ctaggagaaa aaggtagagt cctatcacta agccgcagat   420

atttatttgt gtgtggctag aatgggatgt tttgaatctt ctgttacaac cttgggaacg   480

tggctgttat ttcaatttat gagccagaaa ttttcacatc ccgaaactgc ccagagttcc   540

23

```
accagcctgg gtatagtatt tgttataatc tagtcgtaac agtagttgag ccaaatctga      600

gttgatctga tgattccgaa cactggagag aatcttgaac aggagtgaag actggcggct      660

aaagcccttc acgagaatgc tcagctgggc ggctccacgc tcatccagtg ggcctaggtt      720

ctgactgacc agca                                                        734
```

<210>   19

<211>   2289

<212>   DNA

<213>   Homo sapiens

<400>   19
```
tcgcggccgc gtcracgcgt ggtaggggc ccagagcaag ccgaaggcaa gcacgatggc      60

gctcaccagc cggcccaccc gcgccccgtg ccgcccggag ccccagcggg cgccccgcag     120

ccgtgccagc gtcacgctgt agcagccgag catcagccga aggaagcac gaaagcggtc     180

agagtctcca ggctcaggtg ggcggcggcg tggaccggcg acgggtggca cagctggcat     240

acgcggtccc tccacaggtg gcggtagacg gcggccggga cggcgagcaa cagggcggcc     300

agccagaccg ccagcagcag gcggcgggcc agggccgggc tgcgcagccg aggcgccagg     360

aaggggcggg tgactgcgag gcagcgctgc aggctgagca ggccggtgag cagcacgctt     420

ggcgtacatg ctgagcgcgc acacgtagta caccgccttg cagcccgcct ggcccagcgg     480

ccaggcctgc cggtcaggaa ggccacaaag agcggcgtga gcagcagcac cgcgccgtcg     540

gccagcgcca ggtgcagcac aagcgtggcc gccagcggtc gcccccgtgc aggctgccag     600

cccgccaagc tccacaccac gaagccgttg ccaggcagcc ccagcagcgc cgccagcagc     660

aggaaggctg tgcctgtggc ccgcgaagtc ttccagctca gcagtgtctc gttccctggg     720

ggacggtagc agaccgacat ccttctgggc ctacaggaca cagaaaaaaa gtggggaagc     780

tgggggaccc tacaaggatc cttggcagga aagcagggat tgtgttcatt ttgagggttt     840

cactgtcagt gagagtctca gcttccatgc aactgtccat cacggctgca actgaaatca     900

gagctgggac acagcgcacc agaagctaaa gtcttgatgc catcaaagga catcccctgc     960

cccattcaca yattcacatc tctgtcacgt ccactaatcg gcaaaaggag aaaagtgaga    1020

gaagatgacc taagtgtgac tgcagcaggc agctctggaa aatgaagcca gagcagtgag    1080

ccagcccctc ctccgaccaa ggaggaagga aagagcagcc ccagcacagg agagaaccac    1140

ccagcccaga agttccaggg aaggaactct ccggtccacc atggagtacc tctcagctct    1200

gaaccccagt gacttactca ggtcagtatc taatataagc tcggagtttg gacggagggt    1260

ctggacctca gctccaccac cccagcgacc tttccgtgtc tgtgatcaca agcggaccat    1320

ccggaaaggc ctgacagctg ccacccgcca ggagctgcta gccaaagcat ggagaccct    1380

actgctgaat ggagtgctaa ccctggtgct agaggaggat ggaactgcag tggacagtga    1440

ggacttcttc cagctgctgg aggatgacac gtgcctgatg gtgttgcagt ctggtcagag    1500

ctggagccct acaaggagtg gagtgctgtc atatgggcct ggacgggaga gccccaagca    1560
```

24

```
cagcaaggac atcggccgat tcacctttga cgtgtacaag caaaaccctc gagacctctt    1620

tggcagcctg aatgtcaaag ccacattcta cgggctctac tctatgagtt gtgactttca    1680

aggacttggc ccaaagaaag tactcaggga gctccttcgt tggacctcca cactgctgca    1740

aggcctgggc catatgttgc tgggaatttc ctccaccctt cgtcatgcag tggaggggggc    1800

tgagcagtgg cagcagaagg gccgcctcca ttcctactaa ggggctctga gcttctgccc    1860

ccagaatcat tccaaccgac ccactgcaaa gactatgaca gcatcaaatt tcaggacctg    1920

cagacagtac aggctagata acccacccaa tttccccact gtcctctgat ccccctcgtga    1980

cagaaccttt cagcataacg cctcacatcc caagtctata cccttacctg aagaatgctg    2040

ttctttccta gccacctttc tagcctccca cttgccctga aaggccaaga tcaagatgtc    2100

ccccaggcat cttgatccca gcctgactgc tgctacatct aatcccctac caatgcctcc    2160

tgtccctaaa ctccccagca tactgatgac agccctctct gactttacct tgagatctgt    2220

cttcataccc ttcccctcaa actaacaaaa acatttccaa taaaaatatc aaatatttac    2280

cgtcaaccc                                                             2289
```

<210> 20

<211> 1511

<212> DNA

<213> Homo sapiens

<400> 20

```
cacatttcat cctttttacat ggttcccatc taccctcaca acacatgtca tcaccaaaga     60

cacacataca agctccaatg gcttttgcca ggcaattctt cctccaggac cccatctggc    120

ccctccctca tccctcccct tggactttgc ccttcttact ggccaggcag gggggccaga    180

gtccaggctt gactcattcc caccttgtcc tgggctgaga tcccaggttt gtaacagaaa    240

acaccactaa agccccagca caggagagaa ccacccagcc cagaagttcc agggaaggaa    300

ctctccggtc caccatggag tacctctcag ctctgaaccc cagtgactta ctcaggtcag    360

tatctaatat aagctcggag tttggacgga gggtctggac ctcagctcca ccaccccagc    420

gacctttccg tgtctgtgat cacaagcgga ccatccggaa aggcctgaca gctgccaccc    480

gccaggagct gctagccaaa gcattggaga ccctactgct gaatggagtg ctaaccctgg    540

tgctagagga ggatggaact gcagtggaca gtgaggactt cttccagctg ctggaggatg    600

acacgtgcct gatggtgttg cagtctggtc agagctggag ccctacaagg agtggagtgc    660

tgtcatatgg cctgggacgg gagaggccca agcacagcaa ggacatcgcc cgattcacct    720

ttgacgtgta caagcaaaac cctcgagacc tctttggcag cctgaatgtc aaagccacat    780

tctacgggct ctactctatg agttgtgact tcaaggact tggcccaaag aaagtactca    840

gggagctcct tcgttggacc tccacactgc tgcaaggcct gggccatatg ttgctgggaa    900

tttcctccac ccttcgtcat gcagtggagg gggctgagca gtggcagcag aagggccgcc    960
```

25

```
tccattccta ctaaggggct ctgagcttct gcccccagaa tcattccaac cgacccactg   1020

caaagactat gacagcatca aatttcagga cctgcagaca gtacaggcta gataacccac   1080

ccaatttccc cactgtcctc tgatcccctc gtgacagaac ctttcagcat aacgcctcac   1140

atcccaagtc tataccctta cctgaagaat gctgttcttt cctagccacc tttctggcct   1200

cccacttgcc ctgaaaggcc aagatcaaga tgtcccccag gcatcttgat cccagcctga   1260

ctgctgctac atctaatccc ctaccaatgc ctcctgtccc taaactcccc agcatactga   1320

tgacagccct ctctgacttt accttgagat ctgtcttcat acccttcccc tcaaactaac   1380

aaaaacattt ccaataaaaa tatcaaatat ttaccactaa gacttctgac tccaatttaa   1440

accaggaaag ggatggggtg gatacccccat tttgccctcc cccatcaaca cccagtccca   1500

gatccaaagc c                                                        1511
```

<210> 21

<211> 6530

<212> DNA

<213> Homo sapiens

<400> 21
```
ttttgttagt ttgaggggaa gggtatgaag acagatctca aggtaaagtc agagagggct     60

gtcatcagta tgctggggag tttagggaca ggaggcattg gtaggggatt agatgtagca    120

gcagtcaggc tgggatcaag atgcctgggg gacatcttga tcttggcctt tcagggcaag    180

tgggaggcca gaaaggtggc taggaaagaa cagcattctt caggtaaggg tatagacttg    240

ggatgtgagg cgttatgctg aaaggttctg tcacgagggg atcagaggac agtggggaaa    300

ttgggtgggt tatctagcct gtactgtctg caggtcctga aatttgatgc tgtcatagtc    360

tttgcagtgg gtcggttgga atgattctgg gggcagaagc tcagagcccc ttagtaggaa    420

tggaggcggc ccttctgctg ccactgctca gccccctcca ctgcatgacg aagggtggag    480

gaaattccca gcaacatatg gcccaggcct tgcagcagtg tggaggtcca acgaaggagc    540

tccctgaatg gcagagacaa gaggaaatca gatgatttgg aaaacttggg aggaagccat    600

caagctggga gatgaggact ttccacaagc aagagctaac taggggtagg tgggtgcaag    660

aggacgaatt atggggacta tccaactgta ggggatgggg cagtatgaca tgttgatttc    720

tgacctgagt actttctttg ggccaagtcc ttgaaagtca caactcatag agtagagccc    780

gtagaatgtg gctttgacat tcaggctgcc aaagaggtct cgagggtttt gcttgtacac    840

gtcaaaggtg aatcgggcga tgtccttgct gtgcttgggc ctctcccgtc ccaggccata    900

tgacagcact ccactctgta ggacacccct gtcagtgcag tagatcctca taccagacac    960

ccaccactaa tctccatcag cactgggtca gaccctccct cgcttggact ttctgtccac   1020

tgtgtgacat ccttgacaat tccacaactc ctcctgcacc tggtccccag gatcagggtt   1080

aagctagaga ggaagcccgg gaaagctcta aaggacaggc attggaagca gccccagtat   1140

aggcctctta cccttgtagg gctccagctc tgaccagact gcaacaccat caggcacgtg   1200
```

26

```
tcatcctcca gcagctggaa gaagtcctca ctgtccactg cagttccatc ctcctctagc   1260

accagggtta gcactccatt cagcagtagg gtctccaatg cctgcccaat ggcaagaagc   1320

aagaagggca ggtcttatcc catgcccctt ccctctttag ctgcccaaca tccatcagtt   1380

ggctctagac attggtcgat gtcccacttt gactttccgg cactttgata cctcctaaag   1440

gttgcagctc tccgtgttct tcagtttttg ggggatccta gctagaggct gacctttttc   1500

ctctttgctc ctaccatgtc attggcatct ccccttgctc ccctccaagt cacttctggt   1560

ttggaattgg aaagcaagcc aggttctcac gaagtccacc cttctgtctt atctacaatg   1620

ctgcacctca cttcccacac cctcaagagt tctccagaag tgttttcagt aatagtgttt   1680

aacctttttg agtccttact ctgtgccagg tatgaggact ttacctacat tatcctctta   1740

ctcctttcaa caaccctagg aggtgatgta ttattattgc ctttttatag ttgaagaaac   1800

tgaggttttg gtaggttgaa caacttccca aggtttgaca ggcaggaagt ggcagaatca   1860

gaatttgaac ttgatttgtc acacaaatca cctttccata ctagcttctg aattctgtcc   1920

ctcgaactct ccctatctcc tgctaacccc tgctcccata gaaaagctca ctcggtggaa   1980

aatgaacaaa ttgaccagag ctcattaggc ccactccgct gcttttagcc ctcagaggga   2040

ggggcagctg tgtgacttca gccctctgct ccatcatcac aagttgccac tgttgtggag   2100

ccccttggct acccctgcta taggaaccga ggaacttggc ctacttactt tggctagcag   2160

ctcctggcgg gtggcagctg tcaggccttt ccggatggtc cgcttgtgat cacagacacg   2220

gaaaggtcgc tggggtggtg gagctgaggt ccagaccctc cgtccaaact ccgagcttat   2280

attagatact gacctggtag ttgagaagaa aagtcaagaa ggggcgagga ggggcttggt   2340

gagtgtaaag ggcatgatga gggtagagtg gctagagggc tagggaggga gagatctagg   2400

tttatcgatt agggatgagg gagagaccat ggagtgcagg tgggggcggg tggctcagga   2460

gcttgacaag cccactgtgg agtggggagc aggagaggaa ggggtactgg ttagtctcct   2520

aggggctgag tggagtattg ttgccctgcc tatatcccct aaaggtggag ggtagagcgg   2580

agggttagca gtcacctgag taagtcactg gggttcagag ctgagaggta ctccatggtg   2640

gaccggagag ttccttccct ggaacttctg ggctgggtgg ttctctcctg tgctggggct   2700

ttagtggtgt tttctgttac aaacctggga tctcagccca ggacaaggtg ggaatgagtc   2760

aagcctggac tctggccccc ctgcctggcc agtaagaagg gcaaagtcca aggggaggga   2820

tgagggaggg gccagatggg gtcctggagg aagaattgcc tggcaaaagc cattggagct   2880

tgtatgtgtg tctttggtga tgacatgtgt tgtgagggta gatgggaacc atgtaaaagg   2940

atgaaatgtg acttctggtg ttttttttatt tctatggagg gaatttctgg ggacggtttc   3000

tggctctcag gctctgagaa gctgcagttt atgagtggct ctgtgtgtgc tgccacctac   3060

tggagaagcc ataagctgca gctttaggaa aagggaaccc ggggcagagt gtggggaagt   3120

gggatggcag catggcaggg ctttggaaaa tgagaggtga gactgtgtcc aggaagggtg   3180

taaggagagg atggatcctg atacatggat tcaggatcat tagggtcctg tctgggacac   3240

tggccttcct gcttacctgc tctttccttc ctccttggtc ggaggagggg ctggctcact   3300
```

27

```
gctctggctt cattttccag agctgcctgc tgcagtcaca cttaggtcat cttctctcac   3360

ttttctcctt ttgccgatta gtggacgtga cagagatgtg aatggggcag ggatgtcctt   3420

tgatggcatc aagactttag cttctggtgc gctgtgtccc agctctgatt tcagttgcag   3480

ccgtgatgga cagttgcatg gaagctgaga ctctcactga cagtgaaacc ctcaaatgaa   3540

cacaatccct gctttcctgc caaggatcct tgtagggtcc cccagcttcc ccactttttt   3600

tctgtgtcct gacaaagaaa cacagagtaa cttgattgcc ctgtgacctg gccagttgca   3660

tttcccctgc aggcttgagc ccaagccaga gccttgaaaa ggtattcagg ttgttgccca   3720

aaacactgaa aaaaactggc cctggccctg aaccaaatac cttgaaccct cgtaaactcc   3780

ataccctgac ccccttgttt tggatatacc caggtagaac aactctctct cactgtctgt   3840

tgtgaggata cgctgtagcc cactcattaa gtacattctc ctaataaatg ctttggactg   3900

atcaccctgc cagtcttttg tcttgggcaa tctatacttt tctcagaggt tcccaaggcc   3960

tactgaaggg acttaacata ctcttaatgg ctttcctctc tcttgtttta ccttatgccc   4020

tcacttcctg agttaacctc ccaaatacag gatcacctgt acccaagccc ttagctcaag   4080

aatacaggat cacctgtacc caagccctta gctcaagctc tgctttggaa gaacccaaac   4140

taagacagtg ctcctggtgc cctccccaag caacctcaag ttctggctgt tacttgagca   4200

gaggcctttc ttttcccttc ccccagctct atccatctgc caggcccccc tcaaatctct   4260

tcatttccaa gttttgcttg acttttccaa gaggagaggg ctgcttctta gtatgtccct   4320

actcatcctt tcctttcttg tcttgtatcc tggtgcagcc tggtaatggg gcctcttcat   4380

ggttgtgtgt catgactccc taaccattat gcctccatgc atcccctgtt cctcctggaa   4440

cctagcacca tgccttacat ggaaaagctg tcattgacag cccggtgaga gccctgaggg   4500

tggagtgact ggggcagggc ctgaggcaag aggtgggagg aggtaggagg ccaggggctc   4560

agccggacca ggagactgga aacaggcaag gataaggcag gtgggggact gagttgtttg   4620

ggtcacctct gcaggccaga gagaccaggc aacatacaca ctgcagaagg tgggctggga   4680

ggattggggc cagagctggg ggagggatga gaacagaagc aggaccagga ttcagcagag   4740

tcctcctatt tccttccacc accagggaat cttactgccc cacttcagct tgtgctgttt   4800

cctggcaagg caggctctca catgcctgga cgcctgggtg cgttggtgat gggaaggagc   4860

agggtgaggg aggggcccca ggagaggccc aggatgagcc tcatcttgtc cctccccatt   4920

cttgtcttac cctctgcaaa tgtgataggc acaggacagg agtaggcacc tcgcctactg   4980

ctgcttaacc tttcagcttc tccaggcccc caatcctgct tgctcccagc ttggtaagta   5040

gatctgtgca cgtcccttta cacccccacca tccagttttg cccagatgtg ctagaatggg   5100

gctggacaaa gaaggagggg ccagactaga ggagtggtgg tagagatagt gacagcctgg   5160

ggtgatgact ttatgcctgt ttaccactga gctctgggaa ggaggccagg agtggggcag   5220

gtcaactgac tgggagcagg ggatctgggt tccaagaagg agttgtgttt gaggtggggt   5280

ctgggtcctc gtggaagtca ggactcccag gcagaaaaga ggcaggctgc agggaagtaa   5340

ggaggaggca tggcaccttc tcatcgggca tcacaggtgg ggttttgccc cacccctgaa   5400

cgccctctgt ggcgccttcc acccacctgt aggcccagaa ggatgtcggt ctgctaccgt   5460
```

cccccaggga	acgagacact	gctgagctgg	aagacttcgc	gggccacagg	cacagccttc	5520
ctgctgctgg	cggcgctgct	ggggctgcct	ggcaacggct	tcgtggtgtg	gagcttggcg	5580
ggctggcggc	ctgcacgggg	gcgaccgctg	gcggccacgc	ttgtgctgca	cctggcgctg	5640
gccgacggcg	cggtgctgct	gctcacgccg	ctctttgtgg	ccttcctgac	ccggcaggcc	5700
tggccgctgg	gccaggcggg	ctgcaaggcg	gtgtactacg	tgtgcgcgct	cagcatgtac	5760
gccagcgtgc	tgctcaccgg	cctgctcagc	ctgcagcgct	gcctcgcagt	cacccgcccc	5820
ttcctggcgc	ctcggctgcg	cagcccggcc	ctggcccgcc	gcctgctgct	ggcggtctgg	5880
ctggccgccc	tgttgctcgc	cgtcccggcc	gccgtctacc	gccacctgtg	gagggaccgc	5940
gtatgccagc	tgtgccaccc	gtcgccggtc	cacgccgccg	cccacctgag	cctggagact	6000
ctgaccgctt	tcgtgcttcc	tttcgggctg	atgctcggct	gctacagcgt	gacgctggca	6060
cggctgcggg	gcgccgctg	gggctccggg	cggcacgggg	cgcgggtggg	ccggctggtg	6120
agcgccatcg	tgcttgcctt	cggcttgctc	tgggccccct	accacgcagt	caaccttctg	6180
caggcggtcg	cagcgctggc	tccaccggaa	ggggccttgg	cgaagctggg	cggagccggc	6240
caggcggcgc	gagcgggaac	tacggccttg	gccttcttca	gttctagcgt	caacccggtg	6300
ctctacgtct	tcaccgctgg	agatctgctg	ccccgggcag	gtccccgttt	cctcacgcgg	6360
ctcttcgaag	gctctgggga	ggcccgaggg	ggcggccgct	ctagggaagg	gaccatggag	6420
ctccgaacta	cccctcagct	gaaagtggtg	gggcagggcc	gcggcaatgg	agacccgggg	6480
ggtgggatgg	agaaggacgg	tccggaatgg	gacctttgac	agcagaccct		6530

```
<210>  22
<211>  424
<212>  DNA
<213>  Artificial sequence


<220>
<223>  Probe
<400>  22
```

ggattagatg	tagcagcagt	caggctggga	tcaagatgcc	tgggggacat	cttgatcttg	60
gcctttcagg	gcaagtggga	ggctagaaag	gtggctagga	aagaacagca	ttcttcaggt	120
aagggtatag	acttgggatg	tgaggcgtta	tgctgaaagg	ttctgtcacg	aggggatcag	180
aggacagtgg	ggaaattggg	tgggttatct	agcctgtact	gtctgcaggt	cctgaaattt	240
gatgctgtca	tagtctttgc	agtgggtcgg	ttggaatgat	tctgggggca	gaagctcaga	300
gccccttagt	aggaatggag	gcggcccttc	tgctgccact	gctcagcccc	ctccactgca	360
tgacgaaggg	tggaggaaat	tcccagcaac	atatggccca	ggccttgcag	cagtgtggag	420
gtcc						424

```
<210>  23
```

```
<211>   424

<212>   DNA

<213>   Artificial sequence


<220>

<223>   Probe

<400>   23
ggacctccac actgctgcaa ggcctgggcc atatgttgct gggaatttcc tccacccttc    60

gtcatgcagt ggaggggct gagcagtggc agcagaaggg ccgcctccat tcctactaag   120

gggctctgag cttctgcccc cagaatcatt ccaaccgacc cactgcaaag actatgacag   180

catcaaattt caggacctgc agacagtaca ggctagataa cccacccaat ttccccactg   240

tcctctgatc ccctcgtgac agaacctttc agcataacgc ctcacatccc aagtctatac   300

ccttacctga agaatgctgt tctttcctag ccacctttct agcctcccac ttgccctgaa   360

aggccaagat caagatgtcc cccaggcatc ttgatcccag cctgactgct gctacatcta   420

atcc                                                                424


<210>   24

<211>   7042

<212>   DNA

<213>   Homo sapiens


<400>   24
aagaagaggt agcgagtgga cgtgactgct ctatcccggg caaaagggat agaaccagag    60

gtggggagtc tgggcagtcg gcgaccgcg aagacttgag gtgccgcagc ggcatccgga   120

gtagcgccgg gctccctccg gggtgcagcc gccgtcgggg gaagggcgcc acaggccggg   180

aagacctcct cccttgtgt ccagtagtgg ggtccaccgg agggcggccc gtgggccggg   240

cctcaccgcg gcgctccggg actgtggggt caggctgcgt tgggtggacg cccacctcgc   300

caaccttcgg aggtccctgg gggtcttcgt gcgccccggg gctgcagaga tccaggggag   360

gcgcctgtga ggcccggacc tgccccgggg cgaagggtat gtggcgagac agagccctgc   420

acccctaatt cccggtggaa aactcctgtt gccgtttccc tccaccggcc tggagtctcc   480

cagtcttgtc ccggcagtgc cgccctcccc actaagacct aggcgcaaag gcttggctca   540

tggttgacag ctcagagaga gaaagatctg agggaagatg gatgcaaaag ctcgaaattg   600

tttgcttcaa catagagaag ctctggaaaa ggacatcaag acatcctaca tcatggatca   660

catgattagt gatggatttt taacaatatc agaagaggaa aaagtaagaa atgagcccac   720

tcaacagcaa agagcagcta tgctgattaa aatgatactt aaaaaagata atgattccta   780

cgtatcattc tacaatgctc tactacatga aggatataaa gatcttgctg cccttctcca   840

tgatggcatt cctgttgtct cttcttccag tgtaaggaca gtcctgtgtg aaggtggagt   900
```

```
accacagagg ccagttgttt ttgtcacaag gaagaagctg gtgaatgcaa ttcagcagaa    960

gctctccaaa ttgaaaggtg aaccaggatg ggtcaccata catggaatgg caggctgtgg   1020

gaagtctgta ttagctgcag aagctgttag agatcattcc cttttagaag gttgtttccc   1080

aggggagtg cattgggttt cagttgggaa acaagacaaa tctgggcttc tgatgaaact   1140

gcagaatctt tgcacacggt tggatcagga tgagagtttt tcccagaggc ttccacttaa   1200

tattgaagag gctaaagacc gtctccgcat tctgatgctt cgcaaacacc caaggtctct   1260

cttgatcttg gatgatgttt gggactcttg ggtgttgaaa gcttttgaca gtcagtgtca   1320

gattcttctt acaaccagag acaagagtgt tacagattca gtaatgggtc ctaaatatgt   1380

agtccctgtg gagagttcct taggaaagga aaaaggactt gaaattttat ccttttttgt   1440

taatatgaag aaggcagatt tgccagaaca agctcatagt attataaaag aatgtaaagg   1500

ctctcccctt gtagtatctt taattggtgc acttttacgt gattttccca atcgctggga   1560

gtactacctc aaacagcttc agaataagca gtttaagaga ataaggaaat cttcgtctta   1620

tgattatgag gctctagatg aagccatgtc tataagtgtt gaaatgctca gagaagacat   1680

caaagattat tacacagatc tttccatcct tcagaaggac gttaaggtgc ctacaaaggt   1740

gttatgtatt ctctgggaca tggaaactga agaagttgaa gacatactgc aggagtttgt   1800

aaataagtct ctttttattct gtgatcggaa tggaaagtcg tttcgttatt atttacatga   1860

tcttcaagta gatttttctta cagagaagaa ttgcagccag cttcaggatc tacataagaa   1920

gataatcact cagtttcaga gatatcacca gccgcatact ctttcaccag atcaggaaga   1980

ctgtatgtat tggtacaact ttctggccta tcacatggcc agtgccaaga tgcacaagga   2040

actttgtgct ttaatgtttt ccctggattg gattaaagca aaaacagaac ttgtaggccc   2100

tgctcatctg attcatgaat ttgtggaata cagacatata ctagatgaaa aggattgtgc   2160

agtcagtgag aattttcagg agttttttatc tttaaatgga caccttcttg gacgacagcc   2220

atttcctaat attgtacaac tgggtctctg tgagccggaa acttcagaag tttatcagca   2280

agctaagctg caggccaagc aggaggtcga taatggaatg ctttacctgg aatggataaa   2340

caaaaaaaac atcacgaatc tttcccgctt agttgtccgc ccccacacag atgctgttta   2400

ccatgcctgc ttttctgagg atggtcagag aatagcttct tgtggagctg ataaaacctt   2460

acaggtgttc aaagctgaaa caggagagaa acttctagaa atcaaggctc atgaggatga   2520

agtgctttgt tgtgcattct ctacagatga cagatttata gcaacctgct cagtggataa   2580

aaaagtgaag atttggaatt ctatgactgg ggaactagta cacacctatg atgagcactc   2640

agagcaagtc aattgctgcc atttcaccaa cagtagtcat catcttctct tagccactgg   2700

gtcaagtgac tgcttcctca aactttggga tttgaatcaa aaagaatgtc gaaataccat   2760

gtttggtcat acaaattcag tcaatcactg cagattttca ccagatgata gcttttggc   2820

tagttgttca gctgatggaa ccttaaagct ttgggatgcg acatcagcaa tgagaggaa   2880

aagcattaat gtgaaacagt tcttcctaaa tttggaggac cctcaagagg atatggaagt   2940

gatagtgaag tgttgttcgt ggtctgctga tggtgcaagg ataatggtgg cagcaaaaaa   3000
```

```
taaaatcttt ttgtggaata cagactcacg ttcaaaggtg gctgattgca gaggacattt   3060

aagttgggtt catggtgtga tgttttctcc tgatggatca tcatttttga catcttctga   3120

tgaccagaca atcaggctct gggagacaaa gaaagtatgt aagaactctg ctgtaatgtt   3180

aaagcaagaa gtagatgttg tgtttcaaga aaatgaagtg atggtccttg cagttgacca   3240

tataagacgt ctgcaactca ttaatggaag aacaggtcag attgattatc tgactgaagc   3300

tcaagttagc tgctgttgct taagtccaca tcttcagtac attgcatttg gagatgaaaa   3360

tggagccatt gagattttag aacttgtaaa caatagaatc ttccagtcca ggtttcagca   3420

caagaaaact gtatggcaca tccagttcac agccgatgag aagactctta tttcaagttc   3480

tgatgatgct gaaattcagg tatggaattg gcaattggac aaatgtatct ttctacgagg   3540

ccatcaggaa acagtgaaag actttagact cttgaaaaat tcaagactgc tttcttggtc   3600

atttgatgga acagtgaagg tatggaatat tattactgga aataaagaaa aagactttgt   3660

ctgtcaccag ggtacagtac tttcttgtga catttctcac gatgctacca agttttcatc   3720

tacctctgct gacaagactg caaagatctg gagtttttgat ctccttttgc cacttcatga   3780

attgaggggc cacaacggct gtgtgcgctg ctctgccttc tctgtggaca gtaccctgct   3840

ggcaacggga gatgacaatg gagaaatcag gatatggaat gtctcaaacg gtgagcttct   3900

tcatttgtgt gctccgcttt cagaagaagg agctgctacc catggaggct gggtgactga   3960

cctttgcttt tctccagatg gcaaaatgct tatctctgct ggaggatata ttaagtggtg   4020

gaacgttgtc actggggaat cctcacagac cttctacaca aatggaacca atcttaagaa   4080

aatacacgtg tcccctgact tcaaaacata tgtgactgtg gataatcttg gtattttata   4140

tattttacag actttagaat aaaatagtta agcattaatg tagttgaact ttttaaattt   4200

ttgaattgga aaaaaattct aatgaaaccc tgatatcaac tttttataaa gctcttaatt   4260

gttgtgcagt attgcattca ttacaaaagt gtttgtggtt ggatgaataa tattaatgta   4320

gcttttttccc aaatgaacat acctttaatc ttgttttttca tgatcatcat taacagtttg   4380

tccttaggat gcaaatgaaa atgtgaatac ataccttgtt gtactgttgg taaaattctg   4440

tcttgatgca ttcaaaatgg ttgacataat taatgagaag aatttggaag aaattggtat   4500

tttaatactg tctgtatttta ttactgttat gcaggctgtg cctcagggta gcagtggcct   4560

gctttttgaa ccacacttac cccaaggggg ttttgttctc ctaaatacaa tcttagaggt   4620

tttttgcact ctttaaattt gcttaaaaaa tattgtgtct gtgtgcatag tctgcagcat   4680

ttcctttaat tgactcaata agtgagtctt ggatttagca ggccccccca cctttttttt   4740

ttgttttttgg agacagagtc ttgctttgtt gccaggctgg agtgcagtgg cgcgatctcg   4800

gctcaccaca atcgctgcct cctgggttca agcaattctc ctgcctcagc ctcccgagta   4860

gctgggacta caggtgtgcg cacatgccag ctaattttt gtattttttag tagagacggg   4920

gtttcaccat gttggccggg atggtctcga tctcttgacc tcatgatcta cccgccttgg   4980

cctcccaaag tgctgagatt acaggcgtga gccaccgtgc ctggccaggc cccttctctt   5040

ttaatggaga cagggtcttg cactatcacc caggctggag tgcagtggca taatcatacc   5100

tcattgcagc ctcagactcc tgggttcaag caatcctcct gcctcagcct cccaagtagc   5160
```

32

```
tgagactgca ggcacgagcc accacaccca gctaattttt aagttttctt gtagagacag    5220

ggtctcacta tgttgtctag gctggtcttg aactcttggc ctcaagtaat cctcctgcct    5280

cagcctccca aagtgttggg attgcagata tgagccactg gcctggcctt cagcagttct    5340

ttttgtgaag taaaacttgt atgttggaaa gagtagattt tattggtcta cccttttctc    5400

actgtagctg ctggcagccc tgtgccatat ctggactcta gttgtcagta tctgagttgg    5460

acactattcc tgctccctct tgtttcttac atatcagact tcttacttga atgaaacctg    5520

atctttccta atcctcactt ttttcttttt taaaaagcag tttctccact gctaaatgtt    5580

agtcattgag gtggggccaa ttttaatcat aagccttaat aagatttttc taagaaatgt    5640

gaaatagaac aattttcatc taattccatt tacttttaga tgaatggcat tgtgaatgcc    5700

attcttttaa tgaatttcaa gagaattctc tggttttctg tgtaattcca gatgagtcac    5760

tgtaactcta gaagattaac cttccagcca acctattttc ctttcccttg tctctctcat    5820

cctcttttcc ttccttcttt cctttctctt cttttatctc caaggttaat caggaaaaat    5880

agcttttgac aggggaaaaa actcaataac tagctatttt tgacctcctg atcaggaact    5940

ttagttgaag cgtaaatcta aagaaacatt ttctctgaaa tatattatta agggcaatgg    6000

agataaatta atagtagatg tggttcccag aaaatataat caaaattcaa agattttttt    6060

tgtttctgta actggaacta aatcaaatga ttactagtgt taatagtaga taacttgttt    6120

ttattgttgg tgcatattag tataactgtg gggtaggtcg gggagagggt aagggaatag    6180

atcactcaga tgtattttag ataagctatt tagcctttga tggaatcata aatacagtga    6240

atacaatcct ttgcattgtt aaggaggttt tttgttttta aatggtgggt caaggagcta    6300

gtttacaggc ttactgtgat ttaagcaaat gtgaaaagtg aaaccttaat tttatcaaaa    6360

gaaatttctg taaatggtat gtctccttag aatacccaaa tcataatttt atttgtacac    6420

actgttaggg gctcatctca tgtaggcaga gtataaagta ttaccttttg gaattaaaag    6480

ccactgactg ttataaagta taacaacaca catcaggttt taaaaagcct tgaatggccc    6540

ttgtcttaaa aagaaattag gagccaggtg cggtggcacg tgcctgtagt cccagctcct    6600

tgggaggctg agacaggagg attccttgag ccctggagtt tgagtccagc ctgggtgaca    6660

tagcaagacc ctgtcttaaa agaaaaatgg gaagaaagac aaggtaacat gaagaaagaa    6720

gagataccta gtatgatgga gctgcaaatt tcatggcagt tcatgcagtc ggtcaagagg    6780

aggattttgt tttgtagttt gcagatgagc atttctaaag catttttccct tgctgtattt    6840

ttttgtatta taaattacat tggacttcat atatataatt tttttttaca ttatatgtct    6900

cttgtatgtt ttgaaactct tgtatttatg atatagctta tatgattttt ttgccttggt    6960

atacatttta aaatatgaat ttaaaaaatt tttgtaaaaa taaaattcac aaaattgttt    7020

tgaaaaacaa aaaaaaaaaa aa    7042
```

<210> 25

<211> 3019

<212> DNA

<213>   Artificial sequence

<220>

<223>   Probe

<220>

<221>   misc_feature

<222>   (2846)..(2846)

<223>   Any nucleotide

<400>   25
tttttttttt ttttttttgaa aaacattttt ggattgtttc attctttgct tgtcatttat   60

ctgttgatta gaccactaaa gtgaaggatt caagctaaat acatcaacct ttctatttag   120

gctttatcag ctatatgtaa attcaattct atcaaaattt tctgagtgcc tcctcagtgt   180

gtctctctga tggttcctgc ccggtatggc tggcatgaag aagatccacg gacttgcgaa   240

tgctaacgcg gggcttgggg atgggtttgg agggtttgtt ttcaaagctt tctggaagtg   300

tggaggagtg tcccccttt cttgcttgta gtgctagctg gtaagcgact tcgaatgcct   360

gtcccagggt taggatgatt tcataggcta aattcacatc aaaggcagta aacacatgac   420

agtagtggtg attagacttc aaatcttttg tgatataggc aaatgttgag aggtcttctg   480

ggtcctgggc agcacaggag atattacgaa tttcatgctc agcaattatg ttcttatttg   540

ttgcatcaat aaatttgact cctttatatg agacagaaag aataatagta gggaccttct   600

tcatttgctc tgtagacttc tgacagttag cccgcatttt tgcacaagca tcttgggttg   660

attctgtccc cctaagctct tttatcagca tagaacctaa ataaaaagct ttgtaatcac   720

acgactggaa gataagcttt tctgggtgat gctgccagta ctgtaccggg gtagaggctg   780

tggcttcatt cggaggtcgc aaggtaatgg aaggttctcc ccagtctcct gtctgagcca   840

tctgcctctc cagttttgat cggggaatat catcaaagta gttttcattt cttctcctcc   900

ttgcatcgcc ctgcatgata atgtgaggaa cgtctaggga gccaccagtg gtgtaagtgc   960

tttggctaag tgatggagac aactgaggag gagtgtgatt accactgggt tccctgaggg   1020

tgatggaccg aggggcttc tgtgggggat cgtcgtgcag cctgtctccc agagatgcca   1080

aaatacgttt cctgtggcca atcaaattga ttttttaaaac attaataagt tcaacctccc   1140

agatttttt caacaggtcc atcgaagtgt agccattaat tagaaaggct ttggtgtagt   1200

cgcccagttc aatggaatcc agccactcag ctacagaggt gggatggtag ccatcatgcc   1260

caatgggtct catctttgga aggagctgga ttgcctgtag aattctttgt ctgtgcccag   1320

aattaaggat tccaatttcc aacaaatcct gatcttccat aacattgctt cccataaact   1380

gcacattgtc aaatccatta gccatcaggt ggttctcgta ctgaggtagc ccaatgcttt   1440

ccaaccattg tcccactgtt tggacagggc atctgggtct tgtggtctca ccattcatct   1500

ctttaagttc gttgttgatt ccaacatcta tggaactcat tattttgtca atttcttccc   1560

34

```
attccgatgt gaaggatggt gttctttcag aattcccttt agaactgtgt tcagcagtgg   1620

aagattcact ccagttaact cttgatgttt tctcattgga aggataggca atgagatcag   1680

aatcagattt agagacactt tttgacaaat gcatgtcgat caaggcttta ggcaatgacc   1740

ttattctccc cagagtacag gctctctcca caaatccccc tgcgttcata acccactggt   1800

ccccattccg agatccactc ctggttgatc ttgtgccaac aatggtatgg ttttcgagtt   1860

ggttgctttt tttgtgaaaa attgttctac tgaccacttt gggtttaatt ttctttacca   1920

aaggttctga gttattttct attggagact gcttgaaagg caaaggactg tttgccaaac   1980

tgacttcatc ttgtcctttt tcacattgtt ctcttttccc atagagatga aatggatttt   2040

cagggggactc acaggctgga gaggatccat ggagcaggcc tgcaaattgc ccaggatcat   2100

attcttttgg gggatcattg tcatcctgtc gggagaggtc atctgtatgg ttagttccct   2160

cattcttgac ttctgtggtt cccacagaag aggtgggtgg actagcagga ggactggcag   2220

taaagcttgt gcaccctgta gatgtgttga tttcaaaata ttcttggttg tgattcattc   2280

ggtgaaaatc cagagaagac acaatggatg ttcgctgttt aggctggggt cgaatgactt   2340

ttacaatatt tttgagggca gtatcagggg atggaggtga acaatcaggt gttgggcctg   2400

ttgagctgtt tctatggtta ctagttcctg gagtagtaac tgctacctca gaggcattat   2460

cagttcttgg ggaaggtgcc cttgcaattt ctaaggagca aggtttcttt gtaacagctg   2520

tgtccatgag atcacacaga aagttctcat tttctgaagg aaatgtatcc agagaagcag   2580

atggtacaat ttccatagtg taatttctct tctttggata ggactcctgg gcaagcatgg   2640

ggaagccaag gttcctacat ccattacacg gagttaatgc ttcccaaagt cctgatggcc   2700

cacacgtatt ttcatcatca tcctcttctt ccacttctcc tggtgacaaa ttgattgtag   2760

atgaggttct tacactctgg cttccatttt tcccaagttc ttcctctgaa atcttgctca   2820

aattatctaa gtagtggtct gatatngtgt ggcacaagtc ttcaaacgaa taatcctttt   2880

cttgacagag ttttatttca tccaagagtt ttgataattc tccagtgacg gtttcacttt   2940

tggtcttttg ggaaggagac tcaacaggag atgaaatgtg tgtttcttgt gttgcatctt   3000

cctgtacagg ctcttcgag                                                 3019
```

```
<210>   26
<211>   1752
<212>   DNA
<213>   Artificial sequence


<220>
<223>   Probe
<400>   26
agaacgcaga ccagcccaag ctgacagctt gagtatgcct tcttctgctg cctggttttg     60

ggggctgtat gacgtactgg tcggtagtaa agattaatat gtaagaaatg tggagctagg    120
```

```
atcaagtcat actccacagc ctgcctggca aactatgttt tacttctgac tttgctctct    180

cgctgagaac attaatctgt caagctggcg ggctcctttg atagcaactt tcccagggggc    240

atgatgtggc aatgccacct ctcagcccag gactaccgct attaccccgt ggacggctac    300

tccctgctta aacgcttccc tcttcatcct cttacaggac ccagatgccc tgtccaaaca    360

gtgggacaat ggttggaaag cattgggcta cctcagtacg agaaccacct gatggctaat    420

ggatttgaca atgtgcagtt tatgggaagc aatgttatgg aagatcagga tttgttggaa    480

attggaatcc ttaattctgg gcacagacaa agaattctac aggcaatcca gctccttcca    540

aagatgagac ccattgggca tgatggctac catcccacct ctgtagctga gtggctggat    600

tccattgaac tgggcgacta caccaaagcc tttctaatta atggctacac ttcgatggac    660

ctgttgaaaa aaatctggga ggttgaactt attaatgttt taaaaatcaa tttgattggc    720

cacaggaaac gtattttggc atctctggga gacaggctgc acgacgatcc cccacagaag    780

cccctcggt ccatcaccct caggacagga gactggggag aaccttccat taccttgcga    840

cctccgaatg aagccacagc ctctaccccg gtacagtact ggcagcatca cccagaaaag    900

cttatcttcc agtcgtgtga ttacaaagct ttttatttag gttctatgct gataaaagag    960

cttagggggga cagaatcaac ccaagatgct tgtgcaaaaa tgcgggctaa ctgtcagaag   1020

tctacagagc aaatgaagaa ggtccctact attattcttt ctgtctcata taaaggagtc   1080

aaatttattg atgcaacaaa taagaacata attgctgagc atgaaattcg taatatctcc   1140

tgtgctgccc aggacccaga agacctctca acatttgcct atatcacaaa agatttgaag   1200

tctaatcacc actactgtca tgtgtttact gcctttgatg tgaatttagc ctatgaaatc   1260

atcctaaccc tgggacaggc attcgaagtc gcttaccagc tagcactaca agcaagaaaa   1320

ggggggacact cctccacact tccagaaagc tttgaaaaca aaccctccaa acccatcccc   1380

aagccccgcg ttagcattcg caagtccgtg gatcttcttc atgccagcca taccgggcag   1440

gaaccatcag agagacacac tgaggaggca ctcagaaaat tttgatagaa ttgaatttac   1500

atatagctga taaagcctaa atagaaaggt tgatgtattt agcttgaatc cttcacttta   1560

gtggtctaat caacagataa atgacaagca aagaatgaaa caatccaaaa atgttttca    1620

aaacaatttt gtgaatttta tttttacaaa aatttttaa attcatattt taaaatgtat    1680

accaaggcaa aaaaatcata taagctatat cataaataca agagtttcaa aacatacaag    1740

agacatataa tg                                                        1752
```

<210> 27
<211> 367
<212> DNA
<213> Artificial sequence

<220>

<223> Probe

36

```
<400>  27
ccgcgttagc attcgcaagt ccgtggatct tcttcatgcc agccataccg ggcaggaacc      60

atcagagaga cacactgagg aggcactcag aaaattttga tagaattgaa tttacatata     120

gctgataaag cctaaataga aaggttgatg tatttagctt gaatccttca ctttagtggt     180

ctaatcaaca gataaatgac aagcaaagaa tgaaacaatc caaaaatgtt tttcaaaaca     240

attttgtgaa ttttattttt acaaaaattt tttaaattca tattttaaaa tgtataccaa     300

ggcaaaaaaa tcatataagc tatatcataa atacaagagt ttcaaaacat acaagagaca     360

tataatg                                                               367


<210>  28

<211>  367

<212>  DNA

<213>  Artificial sequence


<220>

<223>  Probe

<400>  28
cattatatgt ctcttgtatg ttttgaaact cttgtattta tgatatagct tatatgattt      60

ttttgccttg gtatacattt taaaatatga atttaaaaaa ttttttgtaaa aataaaattc     120

acaaaattgt tttgaaaaac attttttggat tgtttcattc tttgcttgtc atttatctgt     180

tgattagacc actaaagtga aggattcaag ctaaatacat caacctttct atttaggctt     240

tatcagctat atgtaaattc aattctatca aaattttctg agtgcctcct cagtgtgtct     300

ctctgatggt tcctgcccgg tatggctggc atgaagaaga tccacggact tgcgaatgct     360

aacgcgg                                                               367


<210>  29

<211>  2457

<212>  DNA

<213>  Homo sapiens


<400>  29
cacgcagcag gatggcaagg gctccgcttg gggtcctgct cctcttgggg cttctcggca      60

ggggtgtggg gaagaacgag gaactgcgtc tttatcacca tctcttcaac aactatgacc     120

caggaagccg gccagtgcgg gagcctgagg atactgtcac catcagcctc aaggtcaccc     180

tgacgaatct catctcactg aatgaaaaag aggagactct caccactagc gtctggattg     240

gaatcgattg gcaggattac cgactcaact acagcaagga cgactttggg ggtatagaaa     300

ccctgcgagt cccttcagaa ctcgtgtggc tgccagagat tgtgctggaa aacatattg      360

atggccagtt cggagtggcc tacgacgcca acgtgctcgt ctacgagggc ggctccgtga     420
```

37

```
cgtggctgcc tccggccatc taccgcagcg tctgcgcagt ggaggtcacc tacttcccct   480

tcgattggca gaactgttcg cttattttcc gctctcagac gtacaatgcc gaagaggtgg   540

agttcacttt tgccgtagac aacgacggca agaccatcaa caagatcgac atcgacacag   600

aggcctatac tgagaacggc gagtgggcca tcgacttctg cccggggggtg atccgccgcc   660

accacggtgg cgccaccgac ggcccagggg agactgacgt catctactcg ctcatcatcc   720

gccggaagcc gctcttctac gtcattaaca tcatcgtgcc ctgtgtgctc atctcgggcc   780

tggtgctgct cgcctacttc ctgccggcgc aggccggcgg ccagaaatgc acggtctcca   840

tcaacgtcct gctcgcccag accgtcttct tgttcctcat tgcccagaaa atcccagaga   900

cttctctgag cgtgccgctc ctgggcaggt tccttatttt cgtcatggtg gtcgccacgc   960

tcattgtcat gaattgcgtc atcgtgctca acgtgtccca gcggacgccc accacccacg  1020

ccatgtcccc gcggctgcgc cacgttctcc tggagctgct gccgcgcctc ctgggctccc  1080

cgccgccgcc cgaggccccc cgggccgcct cgcccccaag gcgggcgtcg tcggtgggct  1140

tattgctccg cgcggaggag ctgatactga aaaagccacg gagcgagctc gtgtttgagg  1200

ggcagaggca ccggcagggg acctggacgg ctgccttctg ccagagcctg ggcgccgccg  1260

cccccgaggt ccgctgctgt gtggatgccg tgaacttcgt ggccgagagc acgagagatc  1320

aggaggccac cggcgaggaa gtgtccgact gggtgcgcat ggggaatgcc cttgacaaca  1380

tctgcttctg ggccgctctg gtgctcttca gcgtgggctc cagcctcatc ttcctcgggg  1440

cctacttcaa ccgagtgcct gatctcccct acgcgccgtg tatccagcct tagctcgcac  1500

cgacttcaat ttcccaccca tctccagtag gaaattgatt ttgaaaaagt aggctgccgc  1560

caccacggca ttatgatccc ttccccctgc tgatcaatct gcagtttgtg aacttcacaa  1620

gaatggtgtg tgcccgttcc ctggcgtgtg taggcctggc cgcagtccag gggtcagcag  1680

gaggaaaggg ttcacatagg ctctcaggtg ccagtcttcc agaaagcaag gactgccctt  1740

cattcagcct tgctgacctc ccagccttttc taaggctcag ccccacggga ctctggtggc  1800

tgccagcttg tgagctatct atctatattc atttcatagc caaacaggag accccttttgc  1860

aggacttgca cacagggagg ctgtagccag gaaaccctct tcttccctgg tctggctctg  1920

ctggagcggg tgggaaccaa acaccttcag tgctggtggc cctcaggccc acaggtttaa  1980

ggctgaggct gccctgaccc ttccacagtc atttcttcta ggttttcttg gcccagcact  2040

gcccatccca ccccatgagg ctcactcatt gcagatccca gcccaccctg cccctttctt  2100

ccccaccctg gaggctctct ctgcctagtc tacagtactg acagaaagca aggacatgcg  2160

gcctgcatgg tgggagctgg ttgaattgtc tttattaaca aacaggatat ccaaggccac  2220

tacattgagg aggggggagg ggggaggggag gagaagggtt acttgctgct cacactatat  2280

acagatgcaa gcaaggggcg tggagagtga gggctccctg ctccctccct ccaccgggga  2340

agggcatggg ctagaagagg agaggggggt cgggaatggg gggaatgttt tggctgcggg  2400

gtcccccctc cattccctgg agtttggggg aagggggaatc attaaagtgc tttcaga     2457
```

<210> 30

38

<211> 4863

<212> DNA

<213> Homo sapiens

<400> 30
```
ggagatagcg cctgtcagtc ggtgggtcgg tcctcgcgcc ggccctcccc ctccccggtc      60

tccgggggag gcgcggtgga gtccgccccc ggggttctcc gatggggggag aagcggcgac     120

ggcggcagtg gagtaaccga gccggagcgt gagcggcccc ggtgccccgt tccccacgga     180

ggccatgggc gacccagccc ccgcccgcag cctggacgac atcgacctgt ccgccctgcg     240

ggaccctgct gggatctttg agcttgtgga ggtggtcggc aatggaacct acggacaggt     300

gtacaagggt cggcatgtca agacggggca gctggctgcc atcaaggtca tggatgtcac     360

ggaggacgag gaggaagaga tcaaacagga gatcaacatg ctgaaaaagt actctcacca     420

ccgcaacatc gccacctact acggagcctt catcaagaag agcccccccgg gaaacgatga     480

ccagctctgg ctggtgatgg agttctgtgg tgctggttca gtgactgacc tggtaaagaa     540

cacaaaaggc aacgccctga aggaggactg tatcgcctat atctgcaggg agatcctcag     600

gggtctggcc catctccatg cccacaaggt gatccatcga gacatcaagg ggcagaatgt     660

gctgctgaca gagaatgctg aggtcaagct agtggatttt ggggtgagtg ctcagctgga     720

ccgcaccgtg ggcagacgga acactttcat tgggactccc tactggatgg ctccagaggt     780

catcgcctgt gatgagaacc ctgatgccac ctatgattac aggagtgata tttggtctct     840

aggaatcaca gccatcgaga tggcagaggg agcccccccct ctgtgtgaca tgcaccccat     900

gcgagccctc ttcctcattc ctcggaaccc tccgcccagg ctcaagtcca agaagtggtc     960

taagaagttc attgacttca ttgacacatg tctcatcaag acttacctga gccgcccacc    1020

cacggagcag ctactgaagt ttcccttcat ccgggaccag cccacggagc ggcaggtccg    1080

catccagctt aaggaccaca ttgaccgatc ccggaagaag cggggtgaga agaggagac    1140

agaatatgag tacagcggca gcgaggagga agatgacagc catggagagg aaggagagcc    1200

aagctccatc atgaacgtgc ctggagagtc gactctacgc cgggagtttc tccggctcca    1260

gcaggaaaat aagagcaact cagaggcttt aaaacagcag cagcagctgc agcagcagca    1320

gcagcgagac cccgaggcac acatcaaaca cctgctgcac cagcggcagc ggcgcataga    1380

ggagcagaag gaggagcggc gccgcgtgga ggagcaacag cggcgggagc gggagcagcg    1440

gaagctgcag gagaaggagc agcagcggcg gctggaggac atgcaggctc tgcggcggga    1500

ggaggagcgg cggcaggcgg agcgcgagca ggaatacaag cggaagcagc tggaggagca    1560

gcggcagtca gaacgtctcc agaggcagct gcagcaggag catgcctacc tcaagtccct    1620

gcagcagcag caacagcagc agcagcttca gaaacagcag cagcagcagc tcctgcctgg    1680

ggacaggaag cccctgtacc attatggtcg gggcatgaat cccgctgaca aaccagcctg    1740

ggcccgagag gtagaagaga aacaaggat gaacaagcag cagaactctc ccttggccaa    1800

gagcaagcca ggcagcacgg ggcctgagcc ccccatcccc caggcctccc cagggccccc    1860
```

39

```
aggacccctt tcccagactc ctcctatgca gaggccggtg gagccccagg agggaccgca    1920

caagagcctg gtggcacacc gggtcccact gaagccatat gcagcacctg taccccgatc    1980

ccagtccctg caggaccagc ccacccgaaa cctggctgcc ttcccagcct cccatgaccc    2040

cgaccctgcc atccccgcac ccactgccac gcccagtgcc cgaggagctg tcatccgcca    2100

gaattcagac cccacctctg aaggacctgg ccccagcccg aatcccccag cctgggtccg    2160

cccagataac gaggccccac ccaaggtgcc tcagaggacc tcatctatcg ccactgccct    2220

taacaccagt ggggccggag ggtcccggcc agcccaggca gtccgtgcca gtaaccccga    2280

cctcaggagg agcgaccctg gctgggaacg ctcggacagc gtccttccag cctctcacgg    2340

gcacctcccc caggctggct cactggagcg gaaccgcgtg ggagtctcct ccaaaccgga    2400

cagctcccct gtgctctccc ctgggaataa agccaagccc gacgaccacc gctcacggcc    2460

aggccggccc gcaagctata agcgagcaat tggtgaggac tttgtgttgc tgaaagagcg    2520

gactctggac gaggcccctc ggcctcccaa gaaggccatg gactactcgt cgtccagcga    2580

ggaggtggaa agcagtgagg acgacgagga ggaaggcgaa ggcgggccag cagaggggag    2640

cagagatacc cctggggggcc gcagcgatgg ggatacagac agcgtcagca ccatggtggt    2700

ccacgacgtc gaggagatca ccgggaccca gcccccatac gggggcggca ccatggtggt    2760

ccagcgcacc cctgaagagg agcggaacct gctgcatgct gacagcaatg ggtacacaaa    2820

cctgcctgac gtggtccagc ccagccactc acccaccgag aacagcaaag gccaaagccc    2880

accctcgaag gatgggagtg gtgactacca gtctcgtggg ctggtaaagg cccctggcaa    2940

gagctcgttc acgatgtttg tggatctagg gatctaccag cctggaggca gtggggacag    3000

catccccatc acagccctag tgggtggaga gggcactcgg ctcgaccagc tgcagtacga    3060

cgtgaggaag ggttctgtgg tcaacgtgaa tcccaccaac acccgggccc acagtgagac    3120

ccctgagatc cggaagtaca agaagcgatt caactccgag atcctctgtg cagccctttg    3180

gggggtcaac ctgctggtgg gcacggagaa cgggctgatg ttgctggacc gaagtgggca    3240

gggcaaggtg tatggactca ttgggcggcg acgcttccag cagatggatg tgctggaggg    3300

gctcaacctg ctcatcacca tctcagggaa aaggaacaaa ctgcgggtgt attacctgtc    3360

ctggctccgg aacaagattc tgcacaatga cccagaagtg gagaagaagc agggctggac    3420

caccgtgggg gacatggagg gctgcgggca ctaccgtgtt gtgaaatacg agcggattaa    3480

gttcctggtc atcgccctca gagctccgt ggaggtgtat gcctgggccc ccaaacccta    3540

ccacaaattc atggccttca agtcctttgc cgacctcccc caccgccctc tgctggtcga    3600

cctgacagta gaggaggggc agcggctcaa ggtcatctat ggctccagtg ctggcttcca    3660

tgctgtggat gtcgactcgg ggaacagcta tgacatctac atccctgtgc acatccagag    3720

ccagatcacg ccccatgcca tcatcttcct ccccaacacc gacggcatgg agatgctgct    3780

gtgctacgag gacgagggtg tctacgtcaa cacgtacggg cgcatcatta aggatgtggt    3840

gctgcagtgg ggggagatgc ctacttctgt ggcctacatc tgctccaacc agataatggg    3900

ctggggtgag aaagccattg agatccgctc tgtggagacg ggccacctcg acgggtctt    3960

catgcacaaa cgagctcaga ggctcaagtt cctgtgtgag cggaatgaca aggtgttttt    4020
```

```
tgcctcagtc cgctctgggg gcagcagcca agtttacttc atgactctga accgtaactg    4080

catcatgaac tggtgacggg gccctgggct ggggctgtcc cacactggac ccagctctcc    4140

ccctgcagcc aggcttcccg ggccgcccct ctttcccctc cctgggcttt tgctttact     4200

ggtttgattt cactggagcc tgctgggaac gtgacctctg acccctgatg ctttcgtgat    4260

cacgtgacca tcctcttccc caacatgtcc tcttcccaaa actgtgcctg tccccagctt    4320

ctggggaggg acacagcttc cccttcccag gaattgagtg ggcctagccc ctcccccctt    4380

ttctccattt gagaggagag tgcttggggc ttgaacccct taccccactg ctgctgactg    4440

ggcagggccc tggaccccctt tatttgcacg tcaggggagc cggctccccc cttgaatgta   4500

ccagaccctg gggggggtca ctgggccta gattttggg gggtcaccag ccactccagg      4560

ggcagggacc atttcttcat tttctgaaag cactttaatg attcccctc ccccaaactc     4620

cagggaatgg aggggggacc ccgccagcca aaacattccc cccattcccg accccccctct   4680

cctcttctag cccatgccct tccccggtgg agggagggag cagggagccc tcactctcca    4740

cgccccttgc ttgcatctgt atatagtgtg agcagcaagt aaccccttctc cctccccccc   4800

caccccctcct caatgtagtg gccttggata tcctgtttgt taataaagac aattcaacca   4860

gct                                                                  4863
```

<210> 31

<211> 283

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 31
```
agctggttga attgtctttta ttaacaaaca ggatatccaa ggccactaca ttgaggaggg     60

gggaggggggg agggaggaga agggttactt gctgctcaca ctatatacag atgcaagcaa    120

ggggcgtgga gagtgagggc tccctgctcc ctccctccac cggggaaggg catgggctag    180

aagaggagag gggggtcggg aatggggggga atgttttggc tgcggggtcc ccccctccatt  240

ccctggagtt tgggggaagg ggaatcatta aagtgctttc aga                      283
```

<210> 32

<211> 283

<212> DNA

<213> Artificial sequence

<220>

<223> Probe

<400> 32
tctgaaagca ctttaatgat tcccctcccc ccaaactcca gggaatggag gggggacccc   60

gcagccaaaa cattcccccc attcccgacc cccctctcct cttctagccc atgcccttcc   120

ccggtggagg gagggagcag ggagccctca ctctccacgc cccttgcttg catctgtata   180

tagtgtgagc agcaagtaac ccttctcctc cctcccccct cccccctcct caatgtagtg   240

gccttggata tcctgtttgt taataaagac aattcaacca gct   283

<210> 33

<211> 2714

<212> DNA

<213> Homo sapiens

<400> 33
ggcacagggc gaggttttat acacctgaaa gaagagaatg tcaagacgaa gtagccgttt   60

acaagctaag cagcagcccc agcccagcca gacggaatcc ccccaagaag cccagataat   120

ccaggccaag aagaggaaaa ctacccagga tgtcaaaaaa agaagagagg aggtcaccaa   180

gaaacatcag tatgaaatta ggaattgttg gccacctgta ttatctgggg ggatcagtcc   240

ttgcattatc attgaaacac ctcacaaaga aataggaaca agtgatttct ccagatttac   300

aaattacaga tttaaaaatc tttttattaa tccttcacct ttgcctgatt taagctgggg   360

atgttcaaaa gaagtctggc taaacatgtt aaaaaaggag agcagatatg ttcatgacaa   420

acattttgaa gttctgcatt ctgacttgga accacagatg aggtccatac ttctagactg   480

gcttttagag gtatgtgaag tatacacact tcatagggaa acattttatc ttgcacaaga   540

ctttttttgat agatttatgt tgacacaaaa ggatataaat aaaaatatgc ttcaactcat   600

tggaattacc tcattattca ttgcttccaa acttgaggaa atctatgctc ctaaactcca   660

agagtttgct tacgtcactg atggtgcttg cagtgaagag gatatcttaa ggatggaact   720

cattatatta aaggctttaa aatgggaact ttgtcctgta acaatcatct cctggctaaa   780

tctctttctc caagttgatg ctcttaaaga tgctcctaaa gttcttctac ctcagtattc   840

tcaggaaaca ttcattcaaa tagctcagct tttagatctg tgtattctag ccattgattc   900

attagagttc cagtacagaa tactgactgc tgctgccttg tgccatttta cctccattga   960

agtggttaag aaagcctcag gtttggagtg ggacagtatt tcagaatgtg tagattggat 1020

ggtacctttt gtcaatgtag taaaaagtac tagtccagtg aagctgaaga cttttaagaa 1080

gattcctatg gaagacagac ataatatcca gacacataca aactatttgg ctatgctgga 1140

ggaagtaaat tacataaaca ccttcagaaa agggggacag ttgtcaccag tgtgcaatgg 1200

aggcattatg acaccaccga gagcactga aaaaccacca ggaaaacact aaagaagata   1260

actaagcaaa caagttggaa ttcaccaaga ttgggtagaa ctggtatcac tgaactacta 1320

aagttttaca gaaagtagtg ctgtgattga ttgccctagc caattcacaa gttacactgc 1380

cattctgatt ttaaaactta caattggcac taaagaatac atttaattat ttcctatgtt 1440

42

```
agctgttaaa gaaacagcag gacttgttta caaagatgtc ttcattccca aggttactgg    1500

atagaagcca accacagtct ataccatagc aatgtttttc ctttaatcca gtgttactgt    1560

gtttatcttg ataaactagg aattttgtca ctggagtttt ggactggata agtgctacct    1620

taaagggtat actaagtgat acagtacttt gaatctagtt gttagattct caaaattcct    1680

acactcttga ctagtgcaat ttggttcttg aaaattaaat ttaaacttgt ttacaaaggt    1740

ttagttttgt aataaggtga ctaatttatc tatagctgct atagcaagct attataaaac    1800

ttgaatttct acaaatggtg aaatttaatg tttttttaaac tagtttattt gccttgccat    1860

aacacatttt ttaactaata aggcttagat gaacatggtg ttcaacctgt gctctaaaca    1920

gtgggagtac caaagaaatt ataaacaaga taaatgctgt ggctccttcc taactggggc    1980

tttcttgaca tgtaggttgc ttggtaataa cctttttgta tatcacaatt tgggtgaaaa    2040

acttaagtac cctttcaaac tatttatatg aggaagtcac tttactactc taagatatcc    2100

ctaaggaatt tttttttttta atttagtgtg actaaggctt tatttatgtt tgtgaaactg    2160

ttaaggtcct ttctaaattc ctccattgtg agataaggac agtgtcaaag tgataaagct    2220

taacacttga cctaaacttc tattttctta aggaagaaga gtattaaata tatactgact    2280

cctagaaatc tatttattaa aaaaagacat gaaaacttgc tgtacatagg ctagctattt    2340

ctaaatattt taaattagct tttctaaaaa aaaatccag cctcataaag tagattagaa    2400

aactagattg ctagtttatt ttgttatcag atatgtgaat ctcttctccc tttgaagaaa    2460

ctatacattt attgttacgg tatgaagtct tctgtatagt ttgtttttaa actaatattt    2520

gtttcagtat tttgtctgaa aagaaaacac cactaattgt gtacatatgt attatataaa    2580

cttaaccttt taatactgtt tattttttagc ccattgttta aaaaataaaa gttaaaaaaa    2640

tttaactgct taaaagtaaa gttttgccat tgcttggaga aactttttttt tccttctctg    2700

cgctgccagc tgta                                                      2714
```

<210>   34

<211>   6773

<212>   DNA

<213>   Homo sapiens

<400>   34
```
caagcatgtg atgttcttgt accttcttct gatagtacat ctcaacagtt gactccatat       60

agtcaagtcc atatttgttt gagatctggc aactatcagg aggtaataca gatttttcatt     120

gaagacaact taaccttgag tttacctgtc cagttccgac agtcagtcct aagagaactc      180

tttaagaaag ctcaacaggg aaatgaagct ctagatgaaa tctgtttttaa agtttgtgcc     240

tgtaatacag tccgtgatat actggaaggc agaacaatta gtgttcaatt taaccagcta      300

tttcttagac caaataaaga gaaaatagac tttcttcttg aggtatgttc aagatcagta      360

aatttagaaa aagcttcaga gtctttgaaa ggaaacatgg ctgcttttct aaagaatgtg      420
```

43

```
tgtctggggt tggaagatct gcagtatgtt ttcatgattt cttcacatga gctttccatt    480

acattgttga aagatgaaga acgaaagcta cttgttgatc agatgaggaa gagatccct    540

agagtaaatc tgtgcattaa acctgtaact tcattttatg atatcccagc ttcagcaagt    600

gtcaacattg gtcagttaga gcatcaactt atattgtcag tggatccttg gaggattaga    660

caaattttaa ttgaattaca tggtatgact tcagagcgcc agttctggac agtgtctaat    720

aagtgggaag taccttctgt ctatagtggt gttatcctgg gaattaaaga caatttaaca    780

agagatttgg tttatattct tatggccaaa ggtttgcact gcagtactgt taaggacttt    840

tcccatgcta aacagctctt tgctgcttgt ttggagttgg taacagagtt ctcaccgaag    900

cttcgtcagg tcatgctgaa tgagatgttg cttttggata ttcatacaca cgaagctggg    960

acagggcagg caggagagag accgccatcc gaccttataa gtagagtacg aggctatctg   1020

gaaatgaggc ttcctgatat tcctcttcgt caagttatag ctgaggaatg tgttgccttt   1080

atgttaaact ggagagaaaa tgaatacctt acactccaag ttcctgcatt tttgcttcag   1140

agtaatccat atgtaaagct tggacagctt ttagcagcta catgcaaaga acttccaggc   1200

cctaaagaaa gcagacggac tgccaaagac ctttgggaag ttgttgttca aatctgtagt   1260

gtgtccagtc agcacaaacg aggaaatgat ggcagagtta gtttaataaa acagagggaa   1320

tctacgttag gtatcatgta tcggagtgaa ctgctttctt ttatcaaaaa attacgagaa   1380

ccactcgttt tgactattat tttatcactc tttgtgaaac ttcacaatgt tcgggaggac   1440

attgtgaatg atattacagc tgaacacatt tctatttggc catcttccat tcccaacctc   1500

cagtctgtgg actttgaagc tgtggcaatc acagtgaaag agctagttcg atatacactc   1560

agtataaatc caaataacca ttcttggtta attatccagg cagatatttta ctttgcaacg   1620

aatcagtatt cagcagctct tcactattac ctccaggcag gagctgtgtg ttctgacttc   1680

tttaacaagg ctgtgccccc tgatgtttat acagaccagg taataaaacg aatgataaaa   1740

tgttgttctt tgctgaattg ccacacacag gtggctattt tatgtcagtt cctcagagaa   1800

attgactaca aaacagcgtt taaatctctg caagaacaaa acagtcatga tgctatggac   1860

tcctactacg actacatatg ggatgttacc attttggaat acttgactta tcttcatcat   1920

aaaagaggag aaacagataa aagacaaatt gcaatcaaag ccatcggcca gacagagttg   1980

aatgcaagca atccagaaga agtgttacag ctggcagcgc agagaaggaa aaaaaagttt   2040

ctccaagcaa tggcaaaact ttactttaa gcagttaaat ttttttaact tttattttt   2100

aaacaatggg ctaaagataa acagtattaa aaggttaagt ttatataata catatgtaca   2160

caattagtgg tgttttcttt tcagacaaaa tactgaaaca aatattagtt taaaaacaaa   2220

ctatacagaa gacttcatac cgtaacaata aatgtatagt ttcttcaaag ggagaagaga   2280

ttcacatatc tgataacaaa ataaactagc aatctagttt tctaatctac tttatgaggc   2340

tggatttttt ttttagaaaa gctaatttaa aatatttaga aatagctagc ctatgtacag   2400

caagttttca tgtctttttt taataaatag atttctagga gtcagtatat atttaatact   2460

cttcttcctt aagaaaatag aagtttaggt caagtgttaa gctttatcac tttgacactg   2520

tccttatctc acaatggagg aatttagaaa ggaccttaac agtttcacaa acataaataa   2580
```

44

```
agccttagtc acactaaatt aaaaaaaaaa attccttagg gatatcttag agtagtaaag    2640

tgacttcctc atataaatag tttgaaaggg tacttaagtt tttcacccaa attgtgatat    2700

acaaaaaggt tattaccaag caacctacat gtcaagaaag ccccagttag gaaggagcca    2760

cagcatttat cttgtttata atttctttgg tactcccact gtttagagca caggttgaac    2820

accatgttca tctaagcctt attagttaaa aaatgtgtta tggcaaggca aataaactag    2880

tttaaaaaac attaaatttc accatttgta gaaattcaag ttttataata gcttgctata    2940

gcagctatag ataaattagt caccttatta caaaactaaa cctttgtaaa caagtttaaa    3000

tttaattttc aagaaccaaa ttgcactagt caagagtgta ggaattttga gaatctaaca    3060

actagattca aagtactgta tcacttagta tacccttaa ggtagcactt atccagtcca    3120

aaactccagt gacaaaattc ctagtttatc aagataaaca cagtaacact ggattaaagg    3180

aaaaacattg ctatggtata gactgtggtt ggcttctatc cagtaacctt gggaatgaag    3240

acatctttgt aaacaagtcc tgctgtttct ttaacagcta acataggaaa taattaaatg    3300

tattctttag tgccaattgt aagtttaaa atcagaatgg cagtgtaact tgtgaattgg    3360

ctaggcaat caatcacagc actactttct gtaaaacttt agtagttcag tgataccagt    3420

tctacccaat cttggtgaat tccaacttgt ttgcttagtt atcttcttta gtgttttcct    3480

ggtggttttt cagtgctctt cggtggtgtc ataatgcctc cattgcacac tggtgacaac    3540

tgtccccctt ttctgaaggt gtttatgtaa tttacttcct cctatacatg ggaagaaatc    3600

atgcactgat ttcataaatc aaagtcaaac cagacttctg ggtacttatt tgagattatt    3660

taggcctaat tttaatagtc tttttatgtc tttgcaagtg tgaagggtca tattctgaaa    3720

gtttctgtaa cgttatatat tttttaaact ctttatctag actgttggca ttgatcttga    3780

gacacttcac aaatcttgct ttgatttcaa agtaatttta ttaacttttc tacattttga    3840

aatcagtgtg cccccttagaa cttctcttcc cctgaaactg cctgaaggag tactctattc    3900

ctaccatcag ttttggtgac ttactagatt cagatagcaa agccaaaaaa ctcacaaaaa    3960

aacataccag catagccaaa tagtttgtat gtgtctggat attatgtctg tcttccatag    4020

gaatcttctt aaaagtcttc agcttcactg gactagtact ttttactaca ttgacaaaag    4080

gtaccatcca atctacacat tctgaaatac tgtcccactc caaacctaga tagatagaaa    4140

aaagttagaa aagcatgaag gttgtacatc agaaactatc ttacatatgt ctgatgtact    4200

tgttgctgtt tttgagatat tttaaaagaa accaaatcat aaccaagaag tttagcatgt    4260

caaaacagat tatcactctc aaactattta catgactatg ttgaagggaa aaaggacttc    4320

agaacttctt aaccagtacc ttctacatat gaaattgaaa tggtcaaatc ccaaagaact    4380

cttaaagcag aactataatg ttgattcatt tcaactgtat ttaaattcca tttggtcttt    4440

ttgttgatac acattcagga ttggaaagta cttctaacag aaagataatt actgaacagc    4500

taatttttt tttgccaaag ttttaaatgc atgtttagca gaatgttaaa gttcagagac    4560

tgtagtccca ttagaagttg tgaaaaggta agaagacaac aaatagagag tcttacctga    4620

ggctttctta accacttcaa tggaggtaaa atggcacaag gcagcagcag tcagtattct    4680
```

45

```
gtactggaac tctaatgaat caatggctag aatacacaga tctaaaagct aacaggaaaa    4740

acaaaagtac aagcaattta ggagaaagat gagtactaaa tgtctcttgc taaaacctta    4800

gggatctaga gataaataag ccaccacccg gccaggcgcg gtggctcacg cctgtaatcc    4860

cagcactttg ggaggccgag gcgggtggat cacaaggtca cgagatcgag accatcctgg    4920

ctaacacggt gaaaccccgc ctctactaaa aaatacaaaa aattagctgg gcttggtggc    4980

acgcgcctgt agtcccagct actcgggagg ctgaggcagg agaatggcgt gaacccggga    5040

tgcagagctt gcagtgagca gagatcgcgc cactgcactc cagcctgggc gaaagagcga    5100

gactccgtct caaaaaaaa aaaaacaagc caccaacctg aaggaagtag acaaggaagg    5160

actgttgcaa tacagtgtga catgtactag caggaagggc acctaatcca gattggaaaa    5220

gatagtgatg gcctcaaatt gccataaatg ggtcttaaaa gataagggag ccaggaagag    5280

taggaggcag agaatgttct aggtataggg acattacttg gaactcagtt cacagttcag    5340

aactcctaag gtgaaaaata aataaggagt accttcattt cttatcaaga aagatgaggg    5400

gtggtggcta gaaagaggca tggtctagat tggatcacaa agggtcttta agaagtcaga    5460

attttatagg ctgattcttg aagctactgg aagatttta aatcaaagtt ccattttaag    5520

aaagatacct tagaatgcag tgaagcagac agactagaag aaaacatgtt tattaagcag    5580

tgagattagt taaaaggctg tataatctag gcaataagag ctgaactagt agcagtggaa    5640

tggtatagtg taaaaggggt agatttcaca gatttgagaa gatacttgtg cagtggaatt    5700

aaacttcaat tctctttgtc ctcattggtc cagaaggtag gagaaatggg agaagagctg    5760

ggaattggaa gtgaaatatt actgttatat acctctagaa agtccacatt gtttatcggc    5820

ttatcaaaga tttaccatca ctatcagaag ggtatagctg cctaggacaa tttgggatgc    5880

taggaattct ggatgaaaaa attaagcttt taataaaaag ttttataaaa taaaccaatt    5940

tcagtatact tagtggttat ccaatttgag tattcataat gtgctagatt taagcaccac    6000

tgcccacaaa ttttaaccta ggtgacttaa taattatccc caaatgtctt ccatatgtta    6060

gattttcaca tcccacatag aataagaggg tagattttct tcacttttgt tatatggcag    6120

atacagcagc cttaagatta cttacgagaa gtaagcaaga aagaatggga tctcctcttt    6180

ttttttttt ttaatttttt gagatggagt cttgctctgt tgctcaggct ggagtgcagt    6240

agtgcgatct cggctcactg caacctccac ctcccaggtt ccagcgattc tcctgcctca    6300

gcctcccaag taacatgttg gctaggctgc ctcagccgcc caaactcctg acctcaagtg    6360

atctgcctgc ctgcctcagc cgcccaaagt gctgagatta gagacctgag ccacagtgcc    6420

cggccagatc ctcctcctcc tctacttact tactttgtta aatatgctag cctggaaaag    6480

tttactttga atttatgttc taaaaaattt ttttaacaaa gtaattttaa ttctgatatt    6540

taacttgata ggcactctgt gtatccaaat gtaaagacat catacagaat aattctatgc    6600

cattataaag cttaaacaca actggcgaaa aaaatgcttt tccccatttt atatcaaaaa    6660

gagatacttt agtttggact cctaaagaat gaaagtactc agaaaagtgt aaggactttg    6720

tttttctaga aatattaagc aacataaaca ctggggacag aactttatgc gtc          6773
```

46

<210> 35

<211> 1590

<212> DNA

<213> Mus musculus

<400> 35
ctgagaacca gacatcagga tggcaggggc tctgcttggt gccctgcttc tcctgacact   60

ctttggcaga agccagggaa agaatgaaga gcttagcctg tatcaccatc tcttcgacaa   120

ttatgatcca gaatgccggc cagttaggag acctgaggac actgtcacca tcaccctcaa   180

ggtcacccta accaacctca tctcactgaa cgagaaagaa gaaactctga ccaccagtgt   240

ctggattggc attgactggc acgactatcg gctcaactac agcaaggacg attttgcagg   300

tgtaggaatc ctccgggtcc cttcagaaca tgtatggctg ccagagattg ttctagaaaa   360

caatattgat gggcagtttg gagtggccta cgacagcaat gttctagtct atgagggagg   420

ctatgtgagc tggttgcccc cagccatcta ccgcagcacc tgcgcagtgg aggtcaccta   480

tttccccttt gactggcaga actgctctct catttttcgc tcccagacct acaatgctga   540

ggaggtggag ttcatctttg ccgtggatga cgacggcaat accatcaaca agattgacat   600

tgacacggca gcttttaccg agaatggaga atgggccata gactactgcc caggcatgat   660

tcgccgctat gagggaggtt ccacagaagg tcctggagaa actgacgtca tctatacgct   720

catcatccgc cggaagccgc ttttttacgt cattaacatc attgtgcctt gcgtgctcat   780

ttctggcttg gtgctgctcg cttacttcct gcctgcgcag gctggtggcc agaaatgcac   840

ggtctctatc aacgtcctgc tagcccagac tgtcttcttg tttctaattg cccagaaaat   900

tccagagact tctctgagcg tgccactgct gggcaggtat cttatattcg tcatggtggt   960

tgccacgctc attgtcatga attgcgtcat cgtgctcaac gtatctttga ggacgccaac   1020

gactcatgct acatcccctc ggctgcgcca gattttatta gagctgctgc cgcgtctcct   1080

gggctcgagc ccaccccccag aggatccccg aactgcctca ccagcgaggc gtgcctcgtc   1140

tgtgggcatt ctgctcagag cggaggagct catcttgaaa aagccgcgga gcgaactcgt   1200

gtttgagggt cagaggcatc ggcacggaac ttggaccgca gccctctgcc agaacctggg   1260

tgctgcagcc ccagaaatcc gctgctgtgt ggatgctgtg aactttgtgg ctgagagcac   1320

aagagaccag gaagccactg gagaggaact gtccgactgg gtgcgtatgg ggaaggccct   1380

ggacaatgtc tgttttggg cagctttggt gctcttcagc gttggttcta ctctcatctt   1440

ccttggggggt tacttcaacc aagttcctga tctcccttac ccaccgtgca tccaaccatg   1500

agcctgcact ggcacccacc tctcccccac cccccaagaa agagatttg aaaacaggcc   1560

gctgacaata aatctggttt gtgaacttgc   1590

<210> 36

<211> 2227

<212> DNA

<213> Mus musculus

<400> 36
```
tgtgagcagc aagtagccct tctccctcct gtatcctttc tcaatgtagt ggccttggat    60

atatcccctt tgttaataaa gacaattcaa ccagcttcca ccattttgag atcctactat   120

tgttctctct caatcctgga gagatttgag agttgagaat gcagagggta gaggaaaggc   180

attaggctct gtgaagttac tgtgataata gagacgaagt aaggtggatg aataggccag   240

ggatcagtcc tgacacggta ggaccctttg agaatagttt ttaccagccc cagcagggcc   300

aggccagact tctggcttca gtgtttctat atctgggtct tgtaaaaacc tcattggcta   360

tcaactagat aaacattctt taggttagaa ggagccaaga gcaaaattga accaattgcc   420

tccaagtgcc tgaccaaacc acccacccat cttctacttc cctgaggagt tggacccacc   480

cacatgacca cacaacccct cgggcagttc acaaaccaga tttattgtca gcggcctgtt   540

ttcaaaatct ctttcttggg gggtggggga gaggtgggtg ccagtgcagg ctcatggttg   600

gatgcacggt gggtaaggga gatcaggaac ttggttgaag taaccccccaa ggaagatgag   660

agtagaacca acgctgaaga gcaccaaagc tgcccaaaaa cagacattgt ccagggcctt   720

ccccatacgc acccagtcgg acagttcctg tgagagagag cttagcgagg gaggagcctg   780

gagggcgggg catctagcac tgctccgcct caacctccca acccacctct ccagtggctt   840

cctggtctct tgtgctctca gccacaaagt tcacagcatc cacacagcag cggatttctg   900

gggctgcagc acccaggttc tggcagaggg ctgctgctaa ggcaacagca agcgctaggt   960

cattaaaaga gcgtcctaac ggcgagtgta tgcctttgac ccaagagcag tgcttaccgg  1020

tccaagttcc gtgccgatgc ctctgaccct caaacacgag ttcgctccgc ggcttttttca  1080

agatgagctc ctccgctctg agcagaatgc ccacagacga ggcacgcctc gctggtgagg  1140

cagttcgggg atcctctggg ggtgggctcg agcccaggag acgcggcagc agctctaata  1200

aaatctgcag ccggggcaga gagaggttcc aagcccgctt cccacccctg ggcagtactt  1260

tctccaacca gcgcttacct ggcgcagccg aggggatgta gcatgagtcg ttggcgtcct  1320

caaagatacg ttgagcacga tgacgcaatt catgacaatg agcgtggcaa ccaccatgac  1380

gaatataaga tacctgatat acagaagcct gatgtcacag cacccacaa acaaggcact  1440

agctgccctc tacctcacaa ataccacctc gcacagctgg tggcgttact tcttgatcct  1500

cctcaacgat gccagtattg tcctggccct tctgcatata ccatctgttg cggacatgaa  1560

ggggattccc agcaatttgg acaccctgct gtgggtctac cacttccaca gctccaccga  1620

ggtgagggta ttagaatggc agaatctgga gaggtcccca gctcttcctg ctatggccct  1680

ttccatgtga tcattccact cactacccct gctcctccag gtggccttac agcctccact  1740

tctatcttcc ctggaacttg ctgtggccgc agctcacgaa tatctggtgc aaaggttcag  1800

agagcttaag tcccaggacc ccctggaatc cgacaagtcg cccacccaga aggccaccct  1860

agggctggtg ctaagagaag ctgcagccag catcatgagc tttggagcca ccttgttaga  1920

ggtgctgctc tgggaggctg agggatggga ataaaagggg gagagggcta ggccaacaaa  1980
```

48

```
agcaaggacc tctagcccat atgccccaat gtagatctcg gccctgtggc tgcagcagga    2040

ggtgcagcga ctggacggcg gcaacgactg cccaggccca gccccagaca ctggggatcc    2100

tggtagggcg ctggcccgtg tagccctggc cgcagggcag gggattcggc aagctggaac    2160

ggcagctggc gcaagtgccc ggtacctgat ccaggggggcg tggttgtacc tgtgtggacg    2220

aggtttg                                                              2227
```

<210> 37

<211> 2472

<212> DNA

<213> Homo sapiens

<400> 37
```
agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg     60

gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgacccttc ccagcttccc    120

gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag    180

gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg    240

aggaggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct    300

cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca    360

ggtatcgcag gacttccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg    420

ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt    480

gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct    540

gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga    600

tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct    660

agaccagttc cagcttcttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca    720

aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg gggggttgcc    780

tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catgggggtca tgtataaaca    840

gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa    900

acagggccca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg    960

cattgggga ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga   1020

agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc   1080

ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt   1140

tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac   1200

ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga   1260

acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga   1320

agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg   1380

agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc   1440
```

49

```
agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga   1500

tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa   1560

agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc   1620

tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct   1680

cttcacccca gctgtcctgg aaaaaaatag acaattttaa aaaccaaaca gaatgggact   1740

gtcttctgca agcctaccta caaacaggta caatgttgtt tttaagtact tactgagtgt   1800

gcgccgggtg caagctgagc tgcagcactg ctgggcccta caaatgcagc gcaagcacct   1860

caagtcgaac cagactgatg caatcaagtg gcgcctaaga aatcacatgg cattttttggt   1920

ggataatctt cagtactatc tccaggtaga tgtgttggag tctcagttct cccagctgct   1980

tcatcagatc aattctaccc gagactttga aagcatccga ttggctcatg accacttcct   2040

gagcaatttg ctggctcaat cctttatcct attgaaacct gtgtttcact gcctgaatga   2100

aatcctagat ctctgtcaca gtttttgttc gctggtcagt cagaacctag cccactgga    2160

tgagcgtgga gccgcccagc tgagcattct cgtgaagggc tttagccgcc agtcttcact   2220

cctgttcaag attctctcca gtgttcggaa tcatcagatc aactcagatt tggctcaact   2280

actgttacga ctagattata acaaatacta tacccaggct ggtggaactc tgggcagttt   2340

cgggatgtga aaatttctgg ctcataaatt gaaataacag ccacgttccc aaggttgtaa    2400

cagaagattc aaaacatccc attctagcca cacacaaata aatatctgcg gcttaaaaaa   2460

aaaaaaaaaa aa                                                       2472
```

<210> 38

<211> 4165

<212> DNA

<213> Homo sapiens

<400> 38
```
agcatcgagt cggccttgtt gcctactgga gtctccgcag agcccgggcg ggagtagctg     60

gtggaccccg ttgagctgcc gaacttccgg gactcccccg cgacccctttc ccagcttccc    120

gtccgctccg ccgcagcgat tgtctcggtg ggttgattcg gcacaaaccg cccgacccag    180

gggccggtgc gcgtgtggaa ggggaagcac tcccctcgtg gtcgcctgga ggtgcgctgg    240

aggagggggt gacataacca gggactcgag gtccgccgtg ggaatgatcc acgaactgct    300

cttggctctg agcgggtacc ctgggtccat tttcacctgg aacaagcgga gtggcctgca    360

ggtatcgcag gacttccctt tcctccaccc cagtgagacc agtgtcctga atcgactctg    420

ccggctcggc acagactata ttcgcttcac tgagttcatt gaacagtaca cgggccatgt    480

gcaacagcag gatcaccatc catctcaaca gggccaaggt gggttacatg gaatctacct    540

gcgggccttc tgcacagggc tggattctgt tttgcagcct tatcgccaag cactgcttga    600

tttggaacaa gagttcctgg gtgatcccca tctctccata tcacatgtca actacttcct    660

agaccagttc cagcttctttt ttccctctgt gatggttgta gtagaacaaa ttaaaagtca    720
```

50

```
aaagattcat ggttgtcaaa tcctggaaac agtctacaaa cacagctgtg ggggttgcc    780

tcctgttcga agtgcactgg aaaaaatcct ggccgtttgt catggggtca tgtataaaca   840

gctctcagcc tggatgctcc atggactcct cttggaccag catgaagaat tctttatcaa   900

acaggggcca tcttctggta atgtcagtgc ccagccagaa gaggacgagg aggatctggg   960

cattggggga ctgacaggaa aacaactgag agaactgcag gacttgcgcc tgattgagga   1020

agagaacatg ctggcaccat ctctgaagca gttttcccta cgagtggaga ttttgccatc   1080

ctacattcca gtgagggttg ctgaaaaaat cctatttgtt ggagaatctg tccagatgtt   1140

tgagaatcaa aatgtgaacc tgactagaaa aggatccatt ttgaaaaacc aggaagacac   1200

ttttgctgca gagctgcacc gtctcaagca gcagccactc ttcagcttgg tggactttga   1260

acaggtggtg gatcgcattc gcagcactgt ggctgagcat ctctggaagt tgatggtaga   1320

agaatccgat ttactgggtc agctgaagat cattaaagac ttttaccttc tgggacgtgg   1380

agaactgttt caggccttca ttgacacagc tcaacacatg ttgaaaacac cacccactgc   1440

agtaactgag catgatgtga atgtggcctt tcaacagtca gcacacaagg tattgctaga   1500

tgatgacaac cttctccctc tgttgcactt gacaatcgag tatcacggaa aggagcacaa   1560

agcagatgct actcaggcaa gagaagggcc ttctcgggaa acttctcccc gggaagcccc   1620

tgcatctggc tgggcagccc taggtctttc ctacaaagta cagtggccac tacatattct   1680

cttcacccca gctgtcctgg aaaagtacaa tgttgttttt aagtacttac tgagtgtgcg   1740

ccgggtgcaa gctgagctgc agcactgctg ggccctacaa atgcagcgca agcacctcaa   1800

gtcgaaccag actgatgcaa tcaagtggcg cctaagaaat cacatggcat ttttggtgga   1860

taatcttcag tactatctcc aggtagatgt gttggagtct cagttctccc agctgcttca   1920

tcagatcaat tctacccgag actttgaaag catccgattg gctcatgacc acttcctgag   1980

caatttgctg gctcaatcct ttatcctatt gaaacctgtg tttcactgcc tgaatgaaat   2040

cctagatctc tgtcacagtt tttgtttgct ggtcagtcag aacctaggcc cactggatga   2100

gcgtggagcc gcccagctga gcattctcgt gaagggcttt agccgccagt cttcactcct   2160

gttcaagatt ctctccagtg ttcggaatca tcagatcaac tcagatttgg ctcaactact   2220

gttacgacta gattataaca aatactatac ccaggctggt ggaactctgg gcagtttcgg   2280

gatgtgaaaa tttctggctc ataaattgaa ataacagcca cgttcccaag gttgtaacag   2340

aagattcaaa acatcccatt ctagccacac acaaataaat atctgcggct tagtgatagg   2400

actctacctt ttctcctaga agcagttact gaacatccag gagtacaact ccttcccatc   2460

attcccatgt ggaagggtct ctcccatcaa ggagaacatg tggcatctct gatcctttac   2520

attgagaaca tttgttggat atgttcattt attcaatagt catttattga gcacctacta   2580

cgtaccttgg tactgttcaa gctgtgggag atacagcggt agacaaacaa tatagagcag   2640

aaagttaaat attttatggt tcatatgtga aaaagtaatt atgtttataa atagactaac   2700

tgctggatgt taccaccaag taagaaagca acaggtaaga taggctttct ctctccctat   2760

accaagtaat ttatacctac acagattggg caattctagc taatgaaaat atacttaaaa   2820
```

51

```
gtatttctta ggccgggcat ggtggctcac acctgtaatc ccagcacttt gggaggccga    2880

ggcgggcgga tcacctgaag tcaggagttt gagaccagcc tgaccaacat gatgaaacct    2940

cgattctact aaaaatacaa aaattagcca ggtgtggtgg catgtgcctg taatcccagc    3000

tactcaggag gctgagacag gagaattgct tgaacctggg aagcagacgc tgcagtgagc    3060

tgagattgtg ccattgcatt ccagcctggg caacaagagc gaaattccgt ctcaaaaaaa    3120

aaaaaaaaaa aaaaagtatt attctccaag aaaaaggtcc ttaagaaaaa attgagatca    3180

agttgttaga tttttaaata ctgaagattg caggcccaat tacccatctt acacaaacca    3240

tagggttga  agttatctta atatggccca gccatcactg gtaatcaata ttcatatcag    3300

tgtaagtaaa aagaaatatt cactgaacaa cgccctccaa actgaaaaag aatgcagtgt    3360

tctggcatca ggttatagtc actgcatctg gttttcatca ctacatattc tacacacact    3420

gggaagctct gacaacttat tccctgctat tatcaactaa agatcaccct ttccactgct    3480

gtctctggag caggagctgg caaactatgg cctgctgtct gttttgtac  agtttttactg    3540

aaacacagcc gtgcccattt gtttactcat tgtctatggt tgctttcatg ccctcacagc    3600

aaaggcgagt agttgtgatg gatcaaatgg cccacaaagc ctgaaatatt tactctttga    3660

ccctttacag aaaaaaacct tgttgacccc tgctttagag aatgagaagc catgcaggga    3720

tcagtgatgc cagaggaagg gaaggaactg cttccagcta ttgtgacaat aataataata    3780

ataatattgg gtctttgact agaacgtgta acatttccag gtgttctcac ttgtgcttcc    3840

catgtttatc ttacggaagg tcattccatc aagcttatgg tcactgtccc ttcatggcag    3900

ttggtccttt cgttctccct ttagctctaa gagttgggga gtacccacag gtgagctgtg    3960

atctcagctc agagagagag catgaggtct tttttaactg tcaggaaaca gagctgtgcc    4020

caattccact caacttttgg cacaactgtt aatctgggcc ttcacctacc ttaaactgag    4080

tttctgcaag catagcattt tagacaccct ggaataacct tttgggaatg atgccacaga    4140

ataaagttca ctcttaactt ttcaa                                         4165
```

<210> 39

<211> 27

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 39
ggagagaacc acccagccca gaagttc　　　　　　　　　　　　　　　　27

<210> 40

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 40
aggaatggag gcggcccttc tgc   23

<210> 41

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 41
cggaggagct catcttgaaa aag   23

<210> 42

<211> 24

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 42
gatcaggaac ttggttgaag taac   24

<210> 43

<211> 25

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 43
tgtgagcagc aagtaaccct tctcc   25

<210> 44

<211> 793

<212> DNA

53

```
<213>  Artificial sequence


<220>

<223>  Probe

<400>  44
acagagttga atgcaagcaa tccagaagaa gtgttacagc tggcagcgca gagaaggaaa        60

aaaaagtttc tccaagcaat ggcaaaactt tacttttaag cagttaaatt tttttaactt       120

ttatttttta aacaatgggc taaaaataaa cagtattaaa aggttaagtt tatataatac       180

atatgtacac aattagtggt gttttctttt cagacaaaat actgaaacaa atattagttt       240

aaaaacaaac tatacagaag acttcatacc gtaacaataa atgtatagtt tcttcaaagg       300

gagaagagat tcacatatct gataacaaaa taaactagca atctagtttt ctaatctact       360

ttatgaggct ggattttttt tttagaaaag ctaatttaaa atatttagaa atagctagcc       420

tatgtacagc aagttttcat gtcttttttt aataaataga tttctaggag tcagtatata       480

tttaatactc ttcttcctta agaaaataga agtttaggtc aagtgttaag ctttatcact       540

ttgacactgt ccttatctca caatggagga atttagaaag gaccttaaca gtttcacaaa       600

cataaataaa gccttagtca cactaaatta aaaaaaaaaa ttccttaggg atatcttaga       660

gtagtaaagt gacttcctca tataaatagt ttgaaagggt acttaagttt ttcacccaaa       720

ttgtgatata caaaaaggtt attaccaagc aacctacatg tcaagaaagc cccagttagg       780

aaggagccac agc                                                         793
```